

Copyright (C) 1993 - 2000 Compugen Ltd.

## OM Protein - protein search, using sw model

Run on: June 2, 2002, 18:02:22 ; Search time 13.08 Seconds  
 (without alignments)

679.734 Million cell updates/sec

Title: US-09-886-400-4  
 Perfect score: 1877  
 Sequence: 1 LRALVFHGNLQYAEIPKSEI.....RRLDAFRAINYNDWRENGEP 364

Scoring table: BLOSUM62

Gapext 0.5

Searched:

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
 Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*

2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*

3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*

4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*

5: /cgn2\_6/ptodata/2/iaa/PCUS\_COMB.pep:\*

6: /cgn2\_6/ptodata/2/iaa/batchfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description
1	88.5		84.2	2	US-08-613-220B-4	Sequence 4, Appli
2	141.5		7.5	647	1 US-08-894-212A-8	Sequence 8, Appli
3	141.5		7.5	649	1 US-07-894-112A-2	Sequence 2, Appli
4	104		7.5	650	1 US-07-893-912A-1	Sequence 1, Appli
5	104		5.5	329	1 US-08-270-013A-2	Sequence 2, Appli
6	104		5.5	329	1 US-08-838-418A-2	Sequence 2, Appli
7	88.5		4.7	2089	1 US-08-418-893D-23	Sequence 24, Appli
8	88.5		4.7	2089	1 US-08-418-893D-24	Sequence 2, Appli
9	87		4.6	235	2 US-09-141-135-2	Sequence 1, Appli
10	87		4.6	657	4 US-09-355-416-1	Sequence 2, Appli
11	86.5		4.6	689	1 US-08-248-021A-2	Sequence 2, Appli
12	85.5		4.6	764	4 US-09-235-451-36	Sequence 36, Appli
13	85.5		4.6	788	2 US-08-907-116-6	Sequence 6, Appli
14	83.5		4.4	641	1 US-07-718-535-3	Sequence 3, Appli
15	83.5		4.4	641	1 US-08-161-999-3	Sequence 3, Appli
16	83.5		4.4	1675	4 US-08-487-283A-2	Sequence 2, Appli
17	83		4.4	506	2 US-08-849-480A-5	Sequence 5, Appli
18	83		4.4	636	4 US-09-564-805-237	Sequence 237, Appli
19	82.5		4.4	549	2 US-08-675-279-59	Sequence 59, Appli
20	82		4.4	303	3 US-09-002-298-1	Sequence 1, Appli
21	82		4.4	623	4 US-09-104-068-4	Sequence 4, Appli
22	82		4.4	637	4 US-09-104-068-2	Sequence 2, Appli
23	81.5		4.3	491	1 US-08-489-733-3	Sequence 3, Appli
24	81.5		4.3	491	2 US-08-993-581B-3	Sequence 6, Appli
25	81.5		4.3	806	1 US-08-451-715A-6	Sequence 2, Appli
26	81		4.3	1065	4 US-09-412-545-2	Sequence 2, Appli
27	80.5		4.3	550	1 US-08-484-494-2	Sequence 2, Appli

## DESCRIPTION

RESULT	1	US-08-613-220B-4
SEQUENCE	4,	Application US-08613220B
PATENT	Reidy, John	
PATENT NO.	5955751	
TITLE OF INVENTION	ALPHA-GALACTOSIDASE	
NUMBER OF SEQUENCES	4	
CORRESPONDENCE ADDRESS:		
ADDRESSEE	Fish & Richardson, P.C.	
STREET	4225 Executive Square, Suite 1400	
CITY	La Jolla	
STATE	CA	
COUNTRY	US	
ZIP	92037	
COMPUTER READABLE FORM:		
MEDIUM TYPE:	Diskette	
COMPUTER:	IBM Compatible	
OPERATING SYSTEM:	Windows 95	
SOFTWARE:	FAASSEQ FOR Windows Version 2.0	
CURRENT APPLICATION DATA:		
APPLICATION NUMBER:	US-08-613, 220B	
FILING DATE:	08-MAR-1996	
CLASSIFICATION:	435	
PRIOR APPLICATION DATA:		
APPLICATION NUMBER:		
FILING DATE:		
ATTORNEY/AGENT INFORMATION:		
NAME:	Halle, Ph.D., Lisa A.	
REGISTRATION NUMBER:	38, 347	
REFERENCE/DOCKET NUMBER:	09010/004001	
TELECOMMUNICATION INFORMATION:		
TELEPHONE:	619-678-5070	
TELEFAX:	619-68-5099	
INFORMATION FOR SEQ ID NO: 4 :		
SEQUENCE CHARACTERISTICS:		
LENGTH:	346 amino acids	
TYPE:	amino acid	
TOPOLOGY:	linear	
MOLECULE TYPE:	protein	
FRAGMENT TYPE:	internal	
US-08-613-220B-4		

RESULT	8	Query	Match	Length	DB ID	Description
Result No.	8					
Score	88.5					
Match Length	4.7					
DB ID						
Description						

## SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description
1	88.5					
2	141.5					
3	141.5					
4	104					
5	104					
6	88.5					
7	88.5					
8	88.5					
9	87					
10	87					
11	86.5					
12	85.5					
13	85.5					
14	83.5					
15	83.5					
16	83.5					
17	83					
18	83					
19	82.5					
20	82					
21	82					
22	82					
23	81.5					
24	81.5					
25	81.5					
26	81					
27	80.5					

Query Match	84.2%	Score 1580;	DB 2;	Length 346;
Best Local Similarity	94.5%	Prod. No. 6.7e-159;		
Matches	344;	Conservative	1;	Mismatches 1;
Gaps	18;	Indels	18;	

1 LRALVFGHNLQYAEIPKSEIPKVIKAYIPVIETLILIKEBIPFGUNITGYTLKFLPKDID 60  
 1 LRALVFGHNLQYAEIPKSEIPKVIKAYIPVIETLILIKEBIPFGUNITGYTLKFLPKDID 57  
 Qy 61 LYRGIGASDLIEELIGTSYTHAILPLPSVEAQVORDREVKEELFELSPKGFWLPELAY 120  
 Db 58 LYKGIGASDLIEELIGTSYTHAILPLPSVEAQVORDREVKEELFELSPKGFWLPELA- 114  
 Qy 121 DPTPAILKDGNEYELFADGAMLFSAHNSAIKPKPLYPHLIKAQREKRFYISYLLG 180  
 Db 115 DPTPAILKDGNEYELFAD- EAMLSAHNSAIKPKLYPHLIKAQREKRFYISYLL- 171  
 Qy 181 LRLRKAIAKLYFEKGTVLKAVKDIAVPWVAVTAVMLGIGRLNPCKVASYWEDKD 240  
 Db 172 LRLRKAIAKLYFEKGTVLK- YKDIAVPWVAVTAVML- IGRDLMNPCKVASYWEDK- 228  
 Qy 241 NTLLYGDIEFIFYRDAGYRTAGYRTAGYMSVGEGLYLTTSRAP 300  
 Db 229 NTLLYGDIEFIFYRDAG- RNSVBLLETLDEINSELCE- PSEUKHSGRELYLTSSRA- 285  
 Qy 301 DKSLRIRNDEGENARNLMLSYNMRCGSLALLAENSARGNEPLPERLDAFRAYNDNGE 360  
 Db 286 DKSLRIRNDEGENARNLML- YNMRCGELAFLAENSARGW- PLPERLDAFRAYNDWRG- 342  
 Qy 361 NGEP 364  
 Db 343 NGEP 346

RESULT 2  
 US-07-894-212A-8  
 ; Sequence 8, Application US/07894212A  
 ; GENERAL INFORMATION:  
 ; APPLICANT: ASADA, KIYOZO  
 ; APPLICANT: UEMORI, TAKASHI  
 ; APPLICANT: MUKAI, HIROYUKI  
 ; APPLICANT: KATO, IKUNOSHIN  
 ; APPLICANT: LADERMAN, KENNETH  
 ; APPLICANT: ANFINSEN, CHRISTIAN  
 ; TITLE OF INVENTION: THE ALPHA-AMYLASE GENE  
 ; NUMBER OF SEQUENCES: 8  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: CUSHMAN, DARBY & CUSHMAN  
 ; STREET: 1100 NEW YORK AVENUE, N.W.  
 ; CITY: WASHINGTON, D.C.  
 ; COUNTRY: U.S.A.  
 ; ZIP: 20005

COMPUTER READABLE FORM:  
 MEDIUM TYPE: DISK  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent In Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/07-894,212A  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: KOKULIS, PAUL N.  
 REGISTRATION NUMBER: 16773  
 REFERENCE/DOCKET NUMBER: 95469/C-1195  
 TELEPHONE: (202) 861-3000  
 TELEX: 6714627 CUSH  
 INFORMATION FOR SEQ ID NO: 8:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 647 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-07-894-212A-8

Query Match 7.5% Score 141.5; DB 1; Length 647;  
 Best Local Similarity 22.8%; Pred. No. 3.4e-06;  
 Matches 89; Conservative 60; Mismatches 127; Indels 115; Gaps 23;

Qy 25 EKAYIPVTLIKEKIP--FGUNITGYTLKFLP--KDIIDLVKGIIASDIEELIGTSY 78  
 Db 27 EKCYNPFLELP--EYPNRKVATISGPQEWQDNRPEYIDLRLSVKRCGVEIVVAGF 84  
 Qy 79 THA1PLPLPSRVEAQVQDFREYKE--ELFELSPKGFWLPELAYDTPATLKDNEYYL 136  
 Db 85 YEPVIALSIP--KEERIEQIRLMEWA SIGFARGWLTERRQPELVKTURESGIDYV 141  
 Qy 137 FADGEAMLFSAHNSAIKPKLYPHLIKQRE- KRFRYSYLLGLRELRKA 187  
 Db 142 IVD-----DHFEMSGAGLSKELYWMPYYEDGGVIAVPIDKLR--YLIFPRPVDKV 192  
 Qy 188 IKL----VFEKGTVLKAV- KDIIDAVPVWTAINTAVMLGIGRLNPCKVASYWI---- 236  
 Db 193 LEYLHSLIDGDESKYAVFHDDGEKEGIWGPGTYEWY-----EKWLREFEDR 239  
 Qy 237 --EDKDNILLYGTDIE---FIGYRDIAVYMSVGEGLYEVTDLSELCLP----- 281  
 Db 240 ISDDEXINLMLYTELEYKPRGLVLPASY-----FEM-----SEWSLPAKQARLF 287  
 Qy 282 ---SPLKHSG---RELYLRSSWAPDKSLRFLREWEGCN--ARLNLSYNNMRGELL 330  
 Db 288 VEFVNEJKVGIFIEKYRVTGCGIN--KNF-FVKYKPSNYYMKRMVLMSKLVWN 339  
 Qy 331 AENSDARGWEPLPERLDAFRAYND--WRG 359  
 Db 340 -----PEARKYLRAQCDAYWHG 358

RESULT 3  
 US-07-894-212A-2  
 ; Sequence 2, Application US/07894212A  
 ; General Information:  
 ; APPLICANT: ASADA, KIYOZO  
 ; APPLICANT: UEMORI, TAKASHI  
 ; APPLICANT: MUKAI, HIROYUKI  
 ; APPLICANT: KATO, IKUNOSHIN  
 ; APPLICANT: LADERMAN, KENNETH  
 ; APPLICANT: ANFINSEN, CHRISTIAN  
 ; Title of Invention: THE ALPHA-AMYLASE GENE  
 ; Number of Sequences: 8  
 ; Correspondence Address:  
 ; Addressee: CUSHMAN, DARBY & CUSHMAN  
 ; Street: 1100 New York Avenue, N.W.  
 ; City: Washington, D.C.  
 ; Country: U.S.A.  
 ; Zip: 20005  
 ; Computer Readable Form:  
 ; Addresssee: CUSHMAN, DARBY & CUSHMAN  
 ; Street: 1100 New York Avenue, N.W.  
 ; City: Washington, D.C.  
 ; Country: U.S.A.  
 ; Medium Type: Disk  
 ; Computer: IBM PC compatible  
 ; Operating System: PC-DOS/MS-DOS  
 ; Software: Patent In Release #1.0, Version #1.25  
 ; Current Application Data:  
 ; Application Number: US/07-894,212A  
 ; Classification: 435  
 ; Attorney/Agent Information:  
 ; Name: KOKULIS, PAUL N.  
 ; Registration Number: 16773  
 ; Reference/Docket Number: 95469/C-1195  
 ; Telephone: (202) 861-3000  
 ; Telex: 6714627 CUSH  
 ; Information for Seq Id No: 8:  
 ; Sequence Characteristics:  
 ; Length: 647 amino acids  
 ; Type: amino acid  
 ; Strandedness: single  
 ; Topology: linear  
 ; Molecular Type: protein  
 ; Information for Seq Id No: 2:  
 ; Sequence Characteristics:

LENGTH: 649 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: Linear  
 MOLECULE TYPE: protein  
 US-07-894-212A-2

Query Match Score 141.5; DB 1; Length 649;  
 Best Local Similarity 22.8%; Pred. No. 3.4e-06;  
 Matches 89; Conservative 60; Mismatches 127; Indels 115; Gaps 23;

25 EKAYIPVIETLKEIIP---FGILNITYGLTKLP--KDIDLYKGIGIASDLIELIGTSY 78  
 | :| :|| | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
 29 EKCYWPFLTL--EYPNMKVAAHTSGPLIENLDNRPEYIDLRSVLVRGQEIVVAGF 86  
 | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
 Qy 79 THAILPLPLPSRLSVEAQYQRDRVE - EELFESPRGFWLBDPPIPAILKONGEYVL 136  
 | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
 Db 87 YEPYLASIP--KEDRLEQIRLMKENAKSGFDARGWVLTTERVQPELYKLKESGIDYV 143  
 | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
 Qy 137 FADGEAMLSFAHNSAIKPPIKLY-PHLIAKRE-----KFRYIYSLGLGRELRA 187  
 | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
 Db 144 IVD----DYHPMASGLSKEEYWPYTEDGEVATFEPIDEKLR--YLIPPFPVDKV 194  
 | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
 Qy 188 IKL--VFGCKVTIKAV- KDIAPVYPWVAVNTAVMLGIGRLPLMNPKKYASWI---- 236  
 | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
 Db 195 LEYLHSLTGDESKEYAVPHDDEGKFGIWPCTYEWYV-----EKGWLREFDFR 241  
 | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
 Qy 237 --EDKDNLILYGDIE----FIGYDIAJGYMSVGLLEVIDEINSELCLP----- 281  
 | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
 Db 242 ISSDEKINIMLYTELEYKVKPRGLVYPLIASY-----FEM-----SENSLPAKQARLF 289  
 | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
 Qy 282 ---SELKHSG---RELYLRTSSWAPDKSLRIWREDEGN--ARINMLSYNMRGELLALL 330  
 | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
 Db 290 VEFVNELVKGIFKPFYVFRGGIW--WRG 359  
 | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
 Qy 331 AENS DARGWEPLPERRLDAFRAYND--WRG 359  
 | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
 Db 342 -----PEARKYLLRAQQNDAYWHG 360  
 | :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|  
 RESULT 4  
 US-07-893-928A-1  
 Sequence 1, Application US/07893928A  
 Patent No. 5578479  
 GENERAL INFORMATION:  
 APPLICANT: LADERMAN, KENNETH  
 ADDRESS: ANFINSEN, CHRISTIAN  
 TITLE OF INVENTION: A AMYLASE FROM HYPERTHERMOPHILIC  
 NUMBER OF SEQUENCES: 3  
 CURRENT APPLICATION DATA:  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Tape  
 OPERATING SYSTEM: IBM PC compatible  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 APPLICATION NUMBER: US/07/893-928A  
 FILING DATE:  
 CLASSIFICATION: 435  
 ATTORNEY/AGENT INFORMATION:  
 NAME: KOKULIS, PAUL N.  
 REGISTRATION NUMBER: 16773  
 REFERENCE/DOCKET NUMBER: 95470/C-1197  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202) 861-5000  
 TELEFAX: (202) 822-0944  
 RESULT 5  
 US-08-270-013B-2  
 Sequence 2, Application US/08270013B  
 Patent No. 5686244  
 GENERAL INFORMATION:  
 APPLICANT: Sogabe et al.  
 TITLE OF INVENTION: PROTEIN HAVING HEAT-RESISTANT MALATE  
 NUMBER OF SEQUENCES: 2  
 CURRENT APPLICATION DATA:  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 APPLICATION NUMBER: JP 164701/1993  
 FILING DATE: 02-JUL-1993  
 ATTORNEY/AGENT INFORMATION:

NAME: Green, Robert F.  
 REGISTRATION NUMBER: 27555  
 REFERENCE/DOCKET NUMBER: 62321  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (312) 616-5600  
 TELEX: (312) 616-5700  
 TELEX: (25)3533  
 PRIORITY: 08-08-270-013B-2  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 329 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 329 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-08-338-418-2

Query Match 5 5%; Score 104; DB 1; Length 329;  
 Best Local Similarity 26.0%; Pred. No. 0.011; Gaps 7;  
 Matches 40; Conservative 24; Mismatches 50; Indels 40; Gaps 7;

Query 10 LQYAEIPKSEIPK-----VIEKAYIPVIETLKEEIPPGLNITGYTLKELPKDIDL 61  
 Db 130 MTTYFKEGFPKNRVIGQSVLDTAR---FRTFVAEE----LNIS-----VKDVTF 175

Query 10 LQYAEIPKSEIPK-----VIEKAYIPVIETLKEEIPPGLNITGYTLKELPKDIDL 61  
 Db 130 MTTYFKEGFPKNRVIGQSVLDTAR---FRTFVAEE----LNIS-----VKDVTF 175

Query 62 VKGGIASDLIEITGTSYTHAI-LPLPLSRVEAQVDRREVKEELFELSPKG-FWLPF 117  
 Db 176 VLGGHGDMDMVPARYSYAGGGPLEKLIPKRLDAVERTKGGEIVNLNGNSAYYAPA 235

Query 118 LAYDPPIPAIKDN-----GYEYL 137  
 Db 236 ASLVENVYEAIIKQRRRILPATAYLESEGYEGIY 269

RESULT 7  
 Sequence 23, Application US/08418893D  
 Patent No. 559520

GENERAL INFORMATION:  
 APPLICANT: ROESSLER, PAUL G  
 ATTORNEY: OHLROSSE, JOHN B  
 TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A  
 NUMBER OF SEQUENCES: 25  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY  
 STREET: 1617 Cole Blvd.  
 CITY: Golden  
 STATE: CO  
 COUNTRY: USA  
 ZIP: 80401-3393

COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/418,893D  
 FILING DATE: April 7, 1995  
 CLASSIFICATION: 800  
 ATTORNEY/AGENT INFORMATION:  
 NAME: O'CONNOR, EDNA  
 REGISTRATION NUMBER: 29-252  
 REFERENCE/DOCKET NUMBER: MR1/NREL IR# 92-4BCON

TELECOMMUNICATION INFORMATION:  
 TELEPHONE: 303-231-1000  
 TELEX: 303-231-1098

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:  
 LENGTH: 2089 amino acids  
 TYPE: amino acid  
 STRANDEDNESS: single  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 HYPOTHETICAL: NO  
 ANTI-SENSE: NO

כג טרכו

Query Match 4.7%; Score 88.5; DB 1; Length 2089;  
 Best Local Similarity 19.6%; Pred. No. 9.2; Mismatches 110; Indels 113;  
 Matches 70; Conservative 65; Mismatches 110; Indels 113;

<b>Y</b> <b>9</b> NLQVAEIPKSEPKVIAKAYIPIVETLIK--EPIP----FGLNITGYTLKFLPKDII <b>b</b> <b>1057</b> SFQFDVPAKDK-A-RTVTRGFFSYTDDASKFQAQOLPPELNSFGSSKIAQDASKEGPVNFWI <b>b</b> <b>62</b> VKGTSIASLDL-IEII--GTSYTHAILPLPLPSRVEAQVQDRREVKEEPELSPKGFWI <b>b</b> <b>1115</b> -GALGDGSDISIEDIEKAATFANKDKLNMGLGVRTTALPLFRGKK-----DESYSS <b>b</b> <b>1199</b> A---YDPIIPAIPLKDNGYBYLFADGEAMLFSAHLSNATKPIPLYPHILIAKP-EK <b>b</b> <b>1165</b> SGFKEDPL----- <b>b</b> <b>173</b> -----RYTISYLGLREL--RKAIAKLYFEGKVTLKAVRDIEAPVWVAVNTAVMIL <b>b</b> <b>1196</b> ERIFAVGRNQIYQSGEKTARRNAAQVVF----LRAISHTPOLTTSGARRALLQQ <b>b</b> <b>224</b> LPMLNPKKVASYWIEDKDNLILYG---TDI-----BFIGYRDIAQYRMSEVGLL <b>b</b> <b>1251</b> LERAQANSKVS-VOSSRSLYHLSPQEQS DATAPEEIAKEFEG----- <b>b</b> <b>273</b> EINSELCLSELKHSGRELYRTSWSAPKSLSIWIWRE-EGNR--"LNMLSYNMRCR <b>b</b> <b>1294</b> : <b>b</b> <b>1294</b> KLKSRL-----AQRLTKLRVDE-TETKVRVTQVDEDGSPRVVPVRLVASSMQ	<b>RESULT</b> <b>8</b> <b>S-08-418-893D-24</b> <b>Sequence 24, Application US/08418893D</b> <b>Patent No. 5591920</b> <b>GENERAL INFORMATION:</b> <b>APPLICANT: ROESSLER, PAUL G</b> <b>APPLICANT: OHEORGUE, JOHN B</b> <b>TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A CARBOXYLASE FROM CYCLOTELLA CRYPTICA</b> <b>NUMBER OF SEQUENCES: 25</b> <b>CORRESPONDENCE ADDRESS:</b> <b>ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY</b> <b>STREET: 1617 Cole Blvd.</b> <b>CITY: Golden</b> <b>STATE: CO</b> <b>COUNTRY: USA</b> <b>ZIP: 80401-3393</b> <b>COMPUTER READABLE FORM:</b> <b>MEDIUM TYPE: Floppy disk</b> <b>COMPUTER: IBM PC Compatible</b> <b>OPERATING SYSTEM: PC-DOS/MS-DOS</b> <b>SOFTWARE: Patentin Release #1.0, Version #1.25</b> <b>CURRENT APPLICATION DATA:</b> <b>APPLICATION NUMBER: US/08418,893D</b> <b>CLASSIFICATION: 400</b> <b>CLASSIFICATION DATE: April 7, 1995</b> <b>PRIOR APPLICATION DATA:</b> <b>APPLICATION NUMBER: US 08/104,938</b> <b>FILING DATE: September 14, 1993</b> <b>CLASSIFICATION: 800</b> <b>ATTORNEY/AGENT INFORMATION:</b> <b>NAME: O'CONNOR, EDNA</b> <b>REGISTRATION NUMBER: 29,252</b> <b>TELECOMMUNICATION INFORMATION:</b> <b>TELEPHONE: 303-231-1000</b> <b>TELEFAX: 303-231-1098</b> <b>TELEX:</b> <b>INFORMATION FOR SEQ ID NO: 24:</b> <b>SEQUENCE CHARACTERISTICS:</b> <b>SEQUENCE: 2089 amino acids</b>
--	---

```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
US-08-418-893-D-24

Query Match Score 88.5; DB 1; Length 2069;
Best Local Similarity 19.6%; Pred. No. 9.2; Mismatches 110; Indels 113;
Matches 70; Conservative 65; MisMatches 110; Indels 113;

Qy 9 NLQYAEIPKSEIPKVKTEKAYIPVIETLIK-EEIP----PGLNITGTYTLKFLPKDII
Db 1057 SFGQFADPAKD--RVTQGFPSVQIDASKFAQQLPBLINNSGSKLAGDASSEGQPVNV
Qy 62 VKGIGADDL-IEII--GTSYHAILPLPLSVEADYQRDREVKPEELFELSPKGWLV
Db 1115 -GALSDDISTEDLERATSANKDKLNMGVRTVTALIUPRGK-----DPSYXSF
Qy 119 A---YDPIIPAIILKDNQNEYLFADGEAMLFSAHLSAIPKPKLYPHLRAQR-EKRV
Db 1165 SGKFEDI-----RGRGPPTHHLELGRLEEN
Qy 173 -----RYISYLLGLREL--RKAIAKLYFEGKVTLKAVKDIEAIVPVWVAVNTAVMLG
Db 1196 ERIPAGYRNQVYVGSEKTARNAAQVF-----LRASITPGLTFSGARRALLOG
Qy 224 LPIMNPKKVASHIEDKDNLLYG----TDI-----EFIGYRDINGRMSVEGLLE
Db 1251 LERAQANSKVS-VQSSSRYLHSLSPEQSDATPEEIAFEFEG-----E
Qy 273 ELSNECLCPSEKLKHSGRELYLRTSSWAPDKSLRIWRDE-GNAR--LNMLSYNMRG
Db 1294 KIKRSI-----AQTETKLKYDE- IETKVRVTVDEDGSPPRVYVRLYASSMQC
Db 1295 KIKRSI-----AQTETKLKYDE- IETKVRVTVDEDGSPPRVYVRLYASSMQC

RESULT 9
US-09-141-135-2
; Sequence 2, Application US/09141135
; Patent No. 5981729
; GENERAL INFORMATION:
; APPLICANT: CHUN, Jong Yoon
; TITLE OF INVENTION: Transcription Factor Gene Induced by Water Deficit
; TITLE OF INVENTION: Acid Isolated from Arabidopsis thaliana
; FILE REFERENCE: 1942/31
; CURRENT APPLICATION NUMBER: US/09/141,135
; CURRENT FILING DATE: 1998-08-27
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: SeqPerfect 6.1/Windows
; SEQ ID NO: 2
; LENGTH: 235
; TYPE: PRT
; ORGANISM: Arabidopsis thaliana
US-09-141-135-2

Query Match Score 4.6%; DB 2; Length 235;
Best Local Similarity 21.0%; Pred. No. 0.39; Mismatches 37; Indels 76
Matches 49; Conservative 37; MisMatches 37; Indels 76
; KNNGKREN-----EQIKSMESETLEPRKVKY-----VARELG
Qy 165 KAOREKRFRYISYLLGLRELRKATKLYFEGKVTLKAVKDIEAIVPVWVAVNTAVMLG
Db 276 -LQPRMFIWFQNKRAWKTKOLEKEYNTLR-----ANYNLASQEFIMKKEK
Qy 225 PLMNPKVYASWIEDKD-----NLLGYGTDFIFIGYRDIG ---7YRMSSV
Qy 66 -LQPRMFIWFQNKRAWKTKOLEKEYNTLR-----ANYNLASQEFIMKKEK
Db 269 EVIDELNSEICLPSLKHK---SCRELYIIRTSSWAPD-KSLRWRDGNRNLNMHL

```

Db 117 SEQNRNEEMORPKPEKKHHECCGDDGLASSSTESNGKSEPPGRDGSVLCNDGYNN 176  
 QY 323 -----MRG-----ELALLAENSIDARGNEPLPERDRAFAYNDWGERGE 363  
 Db 177 NIKTEYFRVQGETDHELMNIVEAD-----DSCLTSSENNGGENSD 217

RESULT 10  
 US-09-355-166-1  
 ; Sequence 1, Application US/09355166  
 ; PATENT NO. 616241  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Genencor International, Inc.  
 ; TITLE OF INVENTION: Alpha/Beta Hydrolase-Fold Enzymes  
 ; FILE REFERENCE: GC511-PCT  
 ; CURRENT FILING DATE: 1999-07-20  
 ; NUMBER OF SEQ ID NOS: 21  
 ; SOFTWARE: FastSEQ for Windows Version 3.0  
 ; SEQ ID NO 1  
 ; LENGTH: 657  
 ; TYPE: PRT  
 ; ORGANISM: Bacillus  
 ; US-09-355-166-1

Query Match 4.6%; Score 87; DB 4; Length 657;  
 Best Local Similarity 22.4%; Pred. No. 2.1; Mismatches 106; Indels 88; Gaps 17;  
 Matches 70; Conservative 49; Mismatches 106; Indels 88; Gaps 17;

QY 48 GYTLKFPLK---DIDL---VKGIGASDLIEI-TGTSYT----HAILPLLPLSRVEA 93  
 Db 75 GRTLAFISDREGDAQYLIMSTEGGEARLKLTIDIPVGVSRLPWSDGESILVTISGEGES 134  
 \* 94 QYQRDREVKY----LFLFSPK---GEWPLPELYAPITPAILK-----DNYE 134  
 Db 135 IDDRKEPQDSYTPVEPVQGLSYKRDQKGLTGRAYQVLVSVKS3GEMELTSHKADHDP 194  
 QY 135 YLPADGEAMLFSAHL---NSAIKP-----IKPLYPH---LIKAQREKFRVIS 176  
 Db 195 AEPDGKWLVLFSANLTETDDASKPKHDYYINLESSDLKQVTPHRGSFGSSSFSPDRYLA 254  
 QY 177 YLGGLRELRKA---IKLVFEGRYT-LKAVKDLEAVPYWVAINTAYMLGIGRLPLMNP 229  
 Db 255 LIGNEKEYKNTAKSIAKAWLYDLEQGRLTCEMLDYLHAD--ALIGDSLIGGAECRPI-- 309  
 QY 230 KTKASWEDKDNLLXLTQTDIEFYDIAGRMSTEGGLEVIDENSELCLPSLKHSGR 289  
 Db 310 -----WTKDSOGFYVISTDQGSTGTI-----YYISIEGLVYPI-NLEKE----- 346  
 Qy 290 ELYLRTSSWAPDK 302  
 Db 347 -YINSFSLSPDE 357

RESULT 12  
 US-09-235-451-36  
 ; Sequence 36, Application US/09235451

COUNTRY: United States of America

ZIP: 77210  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: PatentNet Release #1.0, Version #1.3.0  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/248,021A  
 FILING DATE: 24 MAY 1995  
 CLASSIFICATION: 424  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Kitchell, Barbara S.  
 PATENT NO.: 35,928  
 REGISTRATION NUMBER: TAMK:155  
 REFERENCE/DOCKET NUMBER: TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (512) 418-3000  
 TELEFAX: (512) 474-7577  
 INFORMATION FOR SEQ ID NO: 2:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 689 amino acids  
 TYPE: amino acid  
 STRANDEDNESS:  
 TOPOLOGY: linear  
 US-08-248-021A-2

Query Match 4.6%; Score 86.5; DB 1; Length 689;  
 Best Local Similarity 18.8%; Pred. No. 2.5; Mismatches 38; Indels 103; Gaps 11;  
 Matches 51; Conservative 38; Mismatches 79; Indels 103; Gaps 11;

Qy 33 ETLLKEBIPPGLNITYTFLKFLP-----KDLIDLYKGGIASD----- 69  
 Db 264 EAKVNQNPVPSINLNGTTINOSNLAFSNKPWNKYKNETTKYSVLSDRGYSERDLKHA 323  
 Qy 70 -----LIELIGTSYTHAILPLLPLSRVEAQMQRDREVKEL--FELSPK 111  
 Db 324 KKAYTIVYFKNGKGVHLNSNITANLYHAKDVKRLEVTVKTVSKYKAERTVPTLAVN 383  
 Qy 112 GFWLPL-----AYDPT--IPAIILK-DNG-----YEWIFADCEAM 143  
 Db 384 GASNPFLSDLKFTGDSRSVSYSDIKKVKVSVLKHDRGIGERELYAKATYTVHKNGTKK 443  
 Qy 144 LFTSAHNSATKPKPLYPHLIK-----AQREKFRVISY----- 177  
 Db 444 VIT--NUNSNISQNLNLYVDKINIDIVKTAQAKVYKVYPTIAVNGTTTPASKLKL 501  
 Qy 178 --LIGLREIRKAIKVLFEGKVTLKAVKDE 205  
 Db 502 NRQLIGYQDLNKVKVSVLKD--RGINDIE 529

RESULT 12  
 US-09-235-451-36  
 ; Sequence 36, Application US/09235451

GENERAL INFORMATION:  
 APPLICANT: Caterina, Michael J.  
 APPLICANT: Brake, Anthony J.  
 APPLICANT: Julius, David J.  
 APPLICANT: Caterina, Michael J.  
 APPLICANT: Brake, Anthony J.  
 APPLICANT: Julius, David J.

TITLE OF INVENTION: CAPSAICIN RECEPTOR AND CAPSAICIN RELATED POLYPEPTIDES AND USES THEREOF

FILE REFERENCE: 90/6/084CIP  
 CURRENT APPLICATION NUMBER: US/09/235,451  
 PRIOR APPLICATION NUMBER: 999-01-22  
 CURRENT FILING DATE: 1999-01-22  
 PRIOR APPLICATION NUMBER: 60/072,151  
 PRIOR FILING DATE: 1998-01-22  
 PRIOR APPLICATION NUMBER: 08/915,461  
 CURRENT FILING DATE: 1997-08-20  
 NUMBER OF SEQ ID NOS: 48  
 SOFTWARE: FastSEQ for Windows Version 3.0  
 SEQ ID NO 36  
 LENGTH: 764  
 TYPE: PRT

Organism: Homo sapiens  
US-09-451-36

Query Match 4.6%; Score 85.5; DB 4; Length 764;  
Best Local Similarity 22.4%; Pred. No. 3-8; Mismatches 51; Indels 77; Gaps 16;  
Matches 67; Conservative 51; Mismatches 104; Indels 77; Gaps 18;

RESULT 14  
US-07-718-535-3  
Sequence 3, Application US/07718535  
Patent No. 522784

GENERAL INFORMATION:  
APPLICANT: Salyers, Abigail A.,  
Shoemaker, Nadja B.,  
APPLICANT: Nikolic, Mikel P.,  
TITLE OF INVENTION: Method and Materials For  
Introducing DNA Into Prevotella ruminicola  
NUMBER OF SEQUENCES: 5  
SEQUENCE ADDRESS:  
ADDRESSEE: William Brinks Olds Hofer Gilson and Leone  
STREET: P.O. Box 10395  
CITY: Chicago  
STATE: Illinois  
COUNTRY: USA  
ZIP: 60610

COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage  
COMPUTER: IBM XT  
OPERATING SYSTEM: MS-DOS 3.31  
SOFTWARE: WordPerfect 5.1

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/718,535  
FILING DATE: 05-JUN-1991  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Crook, Wannell M.  
REGISTRATION NUMBER: 31,071  
REFERENCE/DOCKET NUMBER: 3617/22  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (312) 321-4200  
TELEFAX: (312) 321-4299

INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 641 amino acids  
TYPE: amino acid  
STRANDEDNESS:  
TOPOLOGY: unknown  
US-07-718-535-3

Query Match 4.4%; Score 83.5; DB 1; Length 641;  
Best Local Similarity 19.5%; Pred. No. 4-6; Mismatches 53; Indels 103; Gaps 16;  
Matches 73; Conservative 53; Mismatches 143; Indels 103; Gaps 16;

QY 24 TEKAYIPVIEELIKEIPFGFLNITYLKGGINSDLI---BIGHSYT 79  
Db 116 LQRQITIIINK----IDRAGVNLELYLDI---KANLSQDVLFMQNYDGSYP 165

QY 80 HAILPLPLPSVEAQYORDREVKEELF--EFLSPKGFWLPELAYDPIPAIKLNGYEYL 136  
Db 166 VCSQTYKEEYFCVNHDDILERYLADEISADYW----NTIALVAKKVPEVL 219

QY 137 FADGEAMLFSAHLNSAIKPI----KPLYPHLKAQREKRFRYISYLLGLRLR 185  
Db 220 --HGSAM-FNIGNEELDADTSFILPPASVSNRSSLYKIEHPKGKRSFL----- 269

QY 186 KATKLVEGKVTKLAVKDIETAVPWWVAINTAVMLIGRIPLMNKVVASWIEDKDNLILY 245  
Db 270 ---KILDGSILRUDVVRINDSEKFKIKNKTINQGR----- 303

QY 246 GTDIEFIGYRDIAGYMSVEGLIVEIDE----INSELCLCPSEKHSRELYLTSSWA 299  
Db 304 EINDEVGANDA-----IVEDMDDFRIGNYIGAEPOLIQSLSHQPHAL--KSSYR 352

QY 300 PDSSLR-----IWREDEGNARLNMLSYNMRGELALLAENSARGWEPLPER---- 345

Query Match 4.6%; Score 85.5; DB 2; Length 788;  
Best Local Similarity 22.1%; Pred. No. 4; Mismatches 42; Indels 108; Gaps 18;  
Matches 73; Conservative 42; Mismatches 108; Indels 107; Gaps 18;

QY 12 YAEPKSELPKVIE-KAYIPVIETLIKEIPFGFLNITYLKGGINSDLI---BIGHSYT 69  
Db 92 YARQD-HVKLREVSOYLEA---DIFPAY-----RYLID--KNLACMD 133

QY 70 LIEFTGTSYTHAILPLPLPSVEAQYORDREVKEELF--LSPKGFWLPELAYDPIIP 125  
SEQ ID NO 6  
Software: FastSEQ for Windows Version 3.0  
Length: 788  
Type: PRT  
Organism: Archaeoglobus lithotrophicus  
US-08-907-166-6

Query Match 4.6%; Score 85.5; DB 2; Length 788;  
Best Local Similarity 22.1%; Pred. No. 4; Mismatches 42; Indels 108; Gaps 18;  
Matches 73; Conservative 42; Mismatches 108; Indels 107; Gaps 18;

QY 126 AILKNGY-EYLFADGEAMLFSAHLNSAIKPKPLYPHL-----IRAQREK- 170  
Db 192 ISIKSGEYEEILNGNEREL---LTFEVKITRDIDDDIIVGYNODSFDWPIKKRAEKL 247

QY 171 -----RFFYIYSTLGLFELRKAIKLV---FEGKVTKLAV 201  
Db 248 RVKIDIGRORSEAIRGGRPKTAGRLNDVLDIAMSLDVKVKJENVAFLGKKI--EL 305

QY 202 KDIAPVWWAVTAVMUGIRLPLMPKVKYASWIE-DKDNILLYGTIDFIGYRDIAGY 260  
Db 306 ADTPAKDITY-----KHWTGDRRESVIKYSR-----QDILHT 336





Result No.	Score	Query	Match	Length	DB ID	Description
1	1494	79.6	364	2	E71144	hypothetical protein PH0368 - Pyrococcus horikoshii
2	169	9.0	529	2	AC2112	hypothetical protein PH0368 - Pyrococcus horikoshii
3	162.5	8.7	529	2	S76831	hypothetical protein PH0368 - Pyrococcus horikoshii
4	161.5	8.6	633	2	G71241	probable alpha-amylase A (EC 3.2.1.1)
5	157.5	8.4	686	1	ALDYAT	uncharacterized co-enzyme A-dependent esterase
6	145.5	7.8	527	2	F97197	alpha-amylase (EC 3.2.1.1)
7	145.5	7.8	655	2	E75206	alpha-amylase (EC 3.2.1.1)
8	141.5	7.5	467	2	B64501	alpha-amylase (EC 3.2.1.1)
9	141.5	7.5	649	2	A49512	hypothetical protein PH0368 - Pyrococcus horikoshii
10	130.5	7.0	560	2	D71011	hypothetical protein PH0368 - Pyrococcus horikoshii
11	128	6.8	923	2	G83826	hypothetical protein PH0368 - Pyrococcus horikoshii
12	124	6.6	3362	2	A75207	hypothetical protein PH0368 - Pyrococcus horikoshii
13	123.5	6.6	744	2	AC1970	hypothetical protein PH0368 - Pyrococcus horikoshii
14	120.5	6.4	324	1	369553	methanol dehydrogenase (EC 1.1.1.1)
15	119	6.3	602	2	C75120	hypothetical protein PH0368 - Pyrococcus horikoshii
16	113	6.0	1069	2	AE1930	hypothetical protein PH0368 - Pyrococcus horikoshii
17	112	6.0	312	2	I40383	hypothetical protein PH0368 - Pyrococcus horikoshii
18	111	5.9	526	2	D71334	hypothetical protein PH0368 - Pyrococcus horikoshii
19	109.5	5.8	447	2	B90250	alpha amylase [imp]
20	106	5.6	526	2	B70859	hypothetical protein PH0368 - Pyrococcus horikoshii
21	105.5	5.6	902	2	E90270	conserved hypothetical protein PH0368 - Pyrococcus horikoshii
22	104	5.5	704	2	G65857	hypothetical protein PH0368 - Pyrococcus horikoshii
23	103.5	5.5	314	2	F84044	malate dehydrogenase (EC 1.1.1.3)
24	101.5	5.4	705	2	F07475	hypothetical protein PH0368 - Pyrococcus horikoshii
25	100.5	5.4	266	2	G66943	probable xylosidase (EC 3.2.1.7)
26	100.5	5.4	619	2	D1361	probable alpha-amylase (EC 3.2.1.1)
27	98.5	5.3	4445	2	T31067	BIR repeat containing protein (EC 3.2.1.1)
28	98.5	5.2	588	2	AE1926	two-component response regulator (EC 3.2.1.1)
29	98	5.2	503	2	T19319	hypothetical protein PH0368 - Pyrococcus horikoshii

GenCore version 4.5	30	98	5.2	764	1	H70414
Copyright (C) 1993 - 2000 Compugen Ltd.	31	97.5	5.2	402	2	T14710
OM protein - protein search, using SW model	32	97.5	5.2	402	2	AB0193
Run on: June 2, 2002, 18:03:43 ; Search time 20.94 Seconds (without alignments)	33	97.5	5.2	402	2	AB0358
1670.321 Million cell updates/sec	34	97.5	5.2	402	2	AB0472
Title: US-09-886-400-4	35	97.5	5.2	402	2	AC0341
Perfect score: 1877	36	97.5	5.2	402	2	AD0002
Sequence: 1 LRALVFGNLQYAEIIPKSBI.....RLDAFRAYNDWRGENGEP 364	37	97.5	5.2	402	2	AE0190
Scoring table: BLOSUM62	38	97.5	5.2	402	2	AE0442
Gapopen 10.0 , Gapext 0.5	39	97.5	5.2	402	2	AE0267
Searched:	283138 seqs, 96089334 residues	40	97.5	5.2	402	2
Total number of hits satisfying chosen parameters:	283138	41	97.5	5.2	402	2
Minimum DB seq length: 0	42	97.5	5.2	402	2	AE0348
Maximum DB seq length: 200000000	43	97.5	5.2	402	2	AG0210
Post-processing: Minimum Match 0%	44	97.5	5.2	402	2	AG0260
Listing first 45 summaries	45	97.5	5.2	402	2	AG0339
Database : PIR71.*	46	97.5	5.2	402	2	AH0193
1: Pir1,*	47	97.5	5.2	402	2	AJ1000
2: Pir2,*	48	97.5	5.2	402	2	AK1000
3: Pir3,*	49	97.5	5.2	402	2	AM1000
4: Pir4,*	50	97.5	5.2	402	2	AN1000

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR71.\*

1: Pir1,\*

2: Pir2,\*

3: Pir3,\*

4: Pir4,\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Score: 8

Query: PH0368

Match: E71144

Length: 1

Query Match: 1 LRALVFGNLQYAEIIPKSBI.....RLDAFRAYNDWRGENGEP 364;

Best Local Similarity: 97.5%; Score: 1494; DB 2; Length: 364;

Mismatches: 0; Indels: 0; Gaps: 0;

Matches: 1 LRALVFGNLQYAEIIPKSBI.....RLDAFRAYNDWRGENGEP 364;

Molecule type: DNA

Residues: 1-164 <KAW>

Cross-references: GB:AP000002; NID:93236129; PID:BA29442.1; PID:g32

Experimental source: strain OT3

Note: this accession replaces an interim accession for a sequence replaced by GenBank

Genetics: PH0368

Superfamily: Pyrococcus horikoshii hypothetical protein PH0368

Conservation: 49; Mismatches: 0;

Qy: 1 LRALVFGNLQYAEIIPKSBI.....RLDAFRAYNDWRGENGEP 364;

Db: 1 MRALVFGNLQYAEIIPKSBI.....RLDAFRAYNDWRGENGEP 364;

Qy: 1 LRGGGIASDLIEJIGSTTHAIPPLPISRVENQVQDRVEKELLELSPKGWEPELAY 120

Db: 61 LKGEGESELIEJIGSTTHAIPPLTLSRTEQIKDRDREIKEFEVSPGFWLPELAY 120

Qy: 1 DP1IBAIIKDNGEYLFDGEAMIFSAHDNSAATKIPPLYPHILIAKEFREKFYYISLLG 180

Db: 61 DP1IBAIIKDNEYFLRGEAMIFSAHDNSAATKIPPLYPHILIAKEFREKFYYISLLG 180

Qy: 1 LIVGGIASDLIEJIGSTTHAIPPLPISRVENQVQDRVEKELLELSPKGWEPELAY 120

Db: 61 LKGEGESELIEJIGSTTHAIPPLTLSRTEQIKDRDREIKEFEVSPGFWLPELAY 120

Qy: 1 D121DP1IBAIIKDNGEYLFDGEAMIFSAHDNSAATKIPPLYPHILIAKEFREKFYYISLLG 180

Db: 61 D121DP1IBAIIKDNEYFLRGEAMIFSAHDNSAATKIPPLYPHILIAKEFREKFYYISLLG 180

Qy: 1 LRELKAIIKLVFEGRTKLAVK DIEA YPVW VAYN TAVMLGIGRLPIMNPKVKA SWE D K D 240

Db: 181 LRELKAIIKLVFEGRTKLAVK DIEA YPVW VAYN TAVMLGIGRLPIMNPKVKA SWE D K D 240

Qy: 1 D5LSLRWREDEGANRANLMSYNNMGEGLAIAEHDARGWEPLPERDAAFRAYNDWRGENGEP 360

Db: 301 D5LSLRWREDEGANRANLMSYNNMGEGLAIAEHDARGWEPLPERDAAFRAYNDWRGENGEP 360

Qy: 1 D5LSLRWREDEGANRANLMSYNNMGEGLAIAEHDARGWEPLPERDAAFRAYNDWRGENGEP 363

#### ALIGNMENTS

1

hypothetical protein PH0368 - Pyrococcus horikoshii

C; Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 20-Jun-2000  
 C; Accession: S76831  
 R; Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asanizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimojo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, Res. 3, 109-136, 1996  
 A; Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp.  
 A; Reference number: S74322; MUID:97061201  
 A; Accession: S76831  
 A; Status: preliminary  
 A; Molecule type: DNA  
 C; Accession: AC2112  
 A; Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120  
 C; Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 11-Jan-2002  
 C; Cross-references: EMBL:D90916; GB:AB001339; NID:91653715; PID:BAA18743.1; PID:g165  
 R; Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritzz, T.; Sasamoto, S.; Watanabe, A.; Iriuchihara, Nakazaki, S.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.; DNA Res. 8, 205-213, 2001  
 A; Title: Complete Genomic Sequence of the Filamentous Nitrogen fixing Cyanobacterium Anabaena sp. Reference number: AB1807; MUID:21595285; PMID:11759840  
 A; Accession: AC2112  
 A; Status: preliminary  
 A; Molecule type: DNA  
 A; Residues: 1-529 <KAN>  
 A; Cross-references: GB:BA000019; PID:BA374149.1; PID:917131542; GSPDB:GN00179  
 A; Experimental source: strain PCC 7120  
 C; Genetics:  
 A; Gene: alr2450  
 C; Superfamily: Pyrococcus horikoshii hypothetical protein PH1386

Query Match Score 8.7%; Score 16.9%; DB 2; Length 529;  
 Best Local Similarity 21.4%; Pred. No. 0 000017; Gaps 19;  
 Matches 101; Conservative 47; Mismatches 140; Indels 183; Gaps 19;

Qy 3 ALVFGNQLQAYAELPKSE-----IPKVIEKAYIPVI--ENLLIKEIIPFFGILNIT----47  
 Db 7 ALVLHAHLPPVRHESDYYLEEWNLYEATETIPLIVFEGIKRDGFDFKTMMSMTPPL 66  
 Qy 48 -----GYTLKLPK-----56

Query Match Score 8.7%; Score 16.2.5%; DB 2; Length 529;  
 Best Local Similarity 21.4%; Pred. No. 0 000017; Gaps 19;  
 Matches 101; Conservative 47; Mismatches 140; Indels 183; Gaps 19;

Qy 57 DILIDVKGGLASDLIEIGTSYTHAILPLPL - SRVEAQVQDRDREYKEELPELSPKGFW 114  
 Db 127 DLVTARKQFDQSNNLIEITCGATGCTYFPFLMKMFQAVWAKIYFACEHYEENFGRSPGIW 186  
 Qy 115 LPELAYDPIITPAIILDKNGYEWLFADGEAMLFSA-----HNS 151  
 Db 187 LPECAYEGVERMLADAGLYFLDVGHGILYARPRPKFETYAPDFTETGVAAFGRDIESS 246  
 Qy 152 -----AIKIKPLVPHLIK-AQREKRFRYIISVLLGFLERIAKLVF---EGKVT 197  
 Db 247 QWVSSOVGYFGDPYREFYKDLGMAEAEYIKPYIMPNGQRKNIGKYHKITSRDGGLS 306  
 Qy 198 LKAVKDIEAVPVWVAINTAV -----MLGI-TGRPLMNPKVVAS-----WI 236  
 Db 307 EWAKYD --PWAKEKAEEHASNEMYNRQQVGHLSGMGRPLVVSPLYDALEFGHWY 362  
 Qy 237 E-----DKDN-----LLYGTDEFI-----GYRDIAFY-----260  
 Db 363 EGWEFLYLERKSWFQDTEPMTHIADYLRGPNPHQCVRPSQSSTWYKGFHEWLNTNA 422  
 Qy 261 -----RMSVEGLLEYDELSELCPSELKHSGREYL-L-RISWWA 299  
 Db 423 WIYPHLHKAERMIELSHREADEEKK ---ALNQAARELLAQSSDW 468

RESULT 4

Db 305 LSDRALYD---EWAKYKA-----EHAANFYNRBQAELHYGIMQRPP 347  
 Qy 160 -----YPHLIKQCR-----EKFRYIISVLLGFLERLKAKLVFEGKVT----197  
 Db 246 SQQWSSEVGPQAAEKFYYDQLGWAEEYIKPYIMPNGQRKTGKIH-KITGRLG 304  
 Qy 198 -LKAVKDIEAVPVWVAINTAVMLGIGRLPMNPKKVASYEDKDNIL--LYGT-----247  
 Db 305 LSDRALYD---EWAKYKA-----EHAANFYNRBQAELHYGIMQRPP 347  
 Qy 248 ---DIEPIGYRDIA-----YRMSVEGLLEYDELSE-----LCLFSELKH 286  
 Db 348 IVSPYDAAELFGHWWYEGPWFIDYLERKSWDQGTYAMTHADLNEPTQQVCRPSQ-----404  
 Qy 287 SGRELILYLRPSSWADKSLSRIWREDDEGA---RLNMLSYNNRGEALLAENSDAFGWEPL 342  
 Db 405 -----SSWYKGFHEYWL-NETNAWIYPLHKAERMI-EIISTL-BPEDDEGLWRAL 452  
 Qy 343 PERRIDAFAIYNDW 357  
 Db 453 NQAARELLAQSSDW 467

A; Accession: G71241  
 A; Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A; Molecule type: DNA  
 A; Residues: 1-633 <KAN>  
 A; Cross-references: GB:AP000001; NID:93236128; PID:BAA29262.1; PID:g3256579  
 A; Experimental source: strain OM3  
 C; Genetics:

RESULT 3

hypothetical protein - Synechocystis sp. (strain PCC 6803)  
 C; Species: Synechocystis sp.  
 A; Variety: PCC 6803

R; Kawarabayasi, Y.; Setwada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sei M.; Ohfuki, Y.; Funashiki, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushiida, N.; Ogu DNA Res. 5, 55-76, 1998  
 A; Title: Complete sequence and gene organization of the genome of a hyper-thermophilic A; Reference number: A1000; MUID:9844137  
 A; Accession: G71241  
 A; Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A; Molecule type: DNA  
 A; Residues: 1-633 <KAN>  
 A; Cross-references: GB:AP000001; NID:93236128; PID:BAA29262.1; PID:g3256579  
 A; Note: this accession replaces an interim accession for a sequence replaced by GenBank C; Genetics:

A;Gene: PH0193	C;Superfamily: Dictyoglomus thermophilum amylase A	Qy 124 IFAILKDNGYEYLADGEAMLFAHLNSAIKPICKPLYPHLIKAKREKRFRYISYLLGLRE 183
Best Local Similarity 23.4%; Pred. No. 0.00027;	Length 633;	Db 132 IWKYIEAGIIVVVVD-DALIFPSVGL-----KEEDLFGY-YLMDEQG 171
Matches 97; Conservative 60; Mismatches 130; Indels 127; Gaps 25;		
Query Match 8.6%; Score 161.5; DB 2;	Length 633;	
Best Local Similarity 23.4%; Pred. No. 0.00027;	Length 633;	Qy 184 LRAKAIKLVFEGKVLKAVKDIEAVPVWAVNTAYMLGIGRLPMLNPKVAVSWI-----ED 238
Matches 97; Conservative 60; Mismatches 130; Indels 127; Gaps 25;		Db 172 YKLA--VFPISSMKLRYL-----IPADPEETITIDKFASED 206
Qy 8 GNQYAEIPKSEPKVIEKAYIPVIETLIKEIP---FGINITYTLKFLPK---DIDL 61		Qy 239 KDNILLYGTDIFIG----YRDLAGYRMSVEGLLEV-IDELNSELCLPS-----E 283
Db 17 GNFYEV----IKRATYKATRPFELT---EBYPMKVAHISGYLVEWLERNPPEYDD 68		Db 207 KSKIALFDDDEKFGLWPDTYRTV-YE---EGWLETFSKIRENELLVTPVNLYTYMQR 261
Qy 62 VKGIGASDLIELIGTSYTHAILPLPLSRVYEAQVQDRREVKEELELVS-----PK 111		Qy 284 LKHSGRELYLTSSW 298
Db 69 LKSIIKGQVELIVAGYEPILVAPI-----EEDRVEQIKLSKGWARKMGYB 117		Db 262 VPKGR-TYLPASY 275
Qy 112 GFMLPELAYDPILPAIKLKDNGEYELADGEAMLFAHLNSA-IKEIKPLYPHLIKAKREK 170		RESULT 6
Db 118 GLWNTTERWEPEVKTLREAGTEVYILD-----DHFEMAGLSKSEELFWPYYTNGSEA 171		F97197
Qy 171 RFPY----ISYGLURBLKAIKLY-----EGKVLTKAVKDLEAVPVWAVNTAYML 219		Accession F97197
Db 172 IVFPVIDEKLRVLIIPFVNTELLEYLHSLADEDESKVAV FHDDEGKEFGAWPGTHTELY- 229		C;Species: Clostridium acetobutylicum
Qy 220 GIGRPLMNPKVKAVSWI-----EDKDNILLYGTDIFIG----YRDAGY-R 261		C;Date: 14-Sep-2001 #text_change 14-Sep-2001 #text_change 30-Sep-2001
Db 230 -----ERGWLKEFFDRISSDDKINMLYS---BYLSKRPKGLVLYLPLIASYFE 274		R;Noiling, J.; Bretnett, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L.
Qy 262 MSYEGL-----LEYTIDEINSELCLPSEKUHSGSELYLRTSSWAPDSLRTRWEDEN 313		; Daly, M.J.; Kononin, E.V.; Smith, D.R.
Db 275 MSENWLSPARQAKLFEFJKKL-KELALFEKRYR----IFYRGQIW--KNE-LKYPGEN 324		J; Bacteriol., 183, 4823-4838, 2001
Qy 314 -- ARNLNSYNMRGELALLAENSARGWEPPLPERRLDAFRAINYD-WRGENG 362		A;Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium
Db 325 YMHRKRMMLS-----KLLRN-----PTARIIFVLRQAQNDAYNHGVFG 362		A;Reference number: A96900; PMID:2135935; PMID:2135925
RESULT 5		A;Accession: F97197
ALDAT		A;Status: preliminary
amylase A (EC 3.2.1.-) - Dictyoglomus thermophilum		A;Molecule type: DNA
C;Species: Dictyoglomus thermophilum		A;Cross references: GB:AE001437; PID:AAK80369; PID:915025430; GSDB:GN00168
C;Accession: S00628; A34969		A;Experimental source: Clostridium acetobutylicum ATCC824
R;PubMed: 174, 15-21, 1988		C;Genetics: CAC2414
A;Title: Cloning and nucleotide sequence of a heat-stable amylase gene from an anaerobic		C;Superfamily: Pyrococcus horikoshii hypothetical protein PH13866
A;Reference number: S00628; MUID:88225097		
A;Accession: S00628		Query Match 7.8%; Score 147; DB 2; Length 527;
A;Molecule type: DNA		Best Local Similarity 18.9%; Pred. No. 0.0026;
A;Residues: 1-686 <FUK>		Matches 92; Conservative 62; Mismatches 120; Indels 212; Gaps 20;
A;Cross references: ENBL:X07896; NID:92688; PID:CAA30735.1; PID:92689		Qy 5 VPHNQLOYAEIPKSE-----IPVIEKAYIPVFE--TLIKEBIPFG-----43
A;Accession: A34969		Db 9 VLHSHMFVRHEPETRSLEERWLWFAEMSECYPLIEVYDNLNDKNIKFRMTMSITPPMLS 68
A;Molecule type: Protein		Qy 44 -----LUNITGYTLKFLPKDII-----DL 61
A;Residues: 2-13 <FUK2>		Db 69 MLDQEYLNRSVLYNLIKTIELSEKELRTKNNRREENKVALFYNKRAENTLKIYKYDNNL 128
C;Superfamily: Dictyoglomus amylase A		Qy 62 VRGGIASD----LIEIIGTSYTHAILPLPLSR--VERAOYQDREYKEELFELSPKGFWLP 116
C;Keywords: glycosidase; hydrolase		Db 129 INAFRKVDRLGVEITCAAHALPLILINQRAVKAQTGQSYINTMGHDFENGIMLP 188
E;2686/Product: amylase A #status: experimental <MAT>		Qy 117 ELAYDPLIPAIKLDNGEYELPFDGEAMLES AHLNSAIKPICKPLYPHLIKAKREKRFRYIS 176
Qy 177 YLGLRLRKAIRKLVEGKVTLKAYDIEAVPVWVAVNTAVMLGIGRPLMNKKVSWI 236		Db 189 ECAYTYGIDNILESERGIKYFISEGKADYASP-----KPMY-----224
Db 225 -----GTNT-----		Qy 177 YLGLRLRKAIRKLVEGKVTLKAYDIEAVPVWVAVNTAVMLGIGRPLMNKKVSWI 236
Query Match 8.4%; Score 157.5; DB 1; Length 686;		Db 225 -----GTNT-----
Best Local Similarity 22.5%; Pred. No. 0.0006;		Qy 237 EDKDNILLYGTDIEFTG-----YRDAGYRMSVBSOLLEYDELNSELCLPSELKH 286
Matches 71; Conservative 53; Mismatches 110; Indels 81; Gaps 15;		Db 241 RDMDS-SYQWSDFNGYPGDFNYREFYDI-GFPLPMYIKPYINE-NGIRIDTGFY 295
Query Match 8.6%; Score 161.5; DB 2;	Length 633;	Qy 287 -----SGRE-LYRFTSS---WAPDKSLRIWREDEGN-ARLMN-----318
Best Local Similarity 23.4%; Pred. No. 0.00027;	Length 633;	Db 296 YKITGNSOERGGYLYNRENAMKKWHERASHFASCRHDOINAAMAMDKPPIIICPYDTELYG 355
Matches 97; Conservative 60; Mismatches 130; Indels 127; Gaps 25;		Qy 319 -----LSYNNRGELALLAENSARGWELPERRDRAFTAYNDW 357

```

RESULT 7
E75206 HWWPEGPDTINAFIRKSADWTSYELITPTPEYLKNNSMVQCSSPSP-----SSW 404
alpha-amylase (or 4-alpha-glucanotransferase) PAB0118 - Pyrococcus abyssi (strain Orsay)
C;Species: Pyrococcus abyssi
C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C;Accession: E75206
R;anonymous, Genoscope
submitted to the ENBL Data Library, July 1999
A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru
A;Descriptor number: A75001
A;Reference number: A75001
A;Accession: E75206
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-655 <KAW>
A;Cross-references: GB:AU248283; GB:AU096836; NID:95457433; PIDN:CA49100_1; PID:9545760
A;Experimental source: strain Orsay
C;Genetics:
A;Gene: amyA; PAB0118
C;Superfamily: Dictyoglomus thermophilum amylase A

Query Match Score 145.5; DB 2; Length 655;
Best: Local Similarity 22.4%; Pred. No. 0.046;
Matches 90; Conservative 55; Missmatches 127; Indels 129; Gaps 22;

QY 25 EKAVTPIEPILIKEIPKIP-----FGINNTGVTKLFLPK--DIDLIVKGTTASDLIEITGSY 78
Db 28 EKAIRPFLEI--EYBPNMKVAHISGLIVLVEENKDYIDLKLKSLVRKGQVEIVYAGF 85
QY 79 THAILPLPLPSRVAQVQDREYKEELPELSPKGFWLPELAYDPPIAIKONGYEYLF 138
Db 86 YEPVILAAIJKEDRLEQIYLKEWKKI-GYDAKSLWLTERNOPELYTLRPAGIEYVV 144
QY 139 DGEAMLFSAHLNSA-IKPIKLPYPLHIAKRE-----KFRFY-----ISYLL 179
Db 145 D-----DHFMSAGLSRQLDNWYYTYDGGEVITVFDIDEKLRLYIFPRYDKVLSYLH 198
QY 180 GLRELRAKAIKLVFEGSKVTLKAVKDIAYPPWVAVNTAYMIGGRIPMLNPKKVASIEDK 239
Db 199 SLASEDESKYAVEH-----DGEKEPFIW-----PM-----TYEWYEK 231
QY 240 D-----NIIYGTDIE-----FIGYRDAGY-----RMSYEG 266
Db 232 GWLREFFDYRSSDEAINMTSEYLQKPKGIVLVPASFYEMSENLSPAQQAKLFEV- 290
QY 267 LLEVDELNSCELPSELEKHSGRLEYLTSSWAPDKSRIWDEGN--ARLNMLSNN 323
Db 291 FVEKELKN-----MFERYEVFGGIW---KNF-FYKYPEANMHKRYMLMS-- 334
QY 324 RGEALLAENSNDARWEPLPERRDAAFRAYND-WRGENG 362
Db 335 ---RILRDNPNSARFF-----VLRQNDAYWHSYFG 362

RESULT 8
B64501 alpha-amylase (EC 3.2.1.1) - Methanococcus jannaschii
C;Species: Methanococcus jannaschii
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: B64501
R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, J.; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Wenslock, K.G.; Merrick, J.M.; Glodek, A.; Sironi, J.D.; Sodow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A;Title: Comparative analysis of the genome of the archaean Methanococcus jannaschii

```

Best Local Similarity 22.8%; Pred. No. 0.009; Matches 89; Conservative 60; Mismatches 127; Indels 115; Gaps 23;	Db 235 RYGNILPA-KTKRSTLRRPF-----LKNGIAVEARNRET-----GIQWSSA 274
Qy 25 EKAYTIVETIILKEIPP--FGLNTGTYLKFIP--ADIDILVAKGGTASDLEIEITGSY 78	RESULT 11 C83826 hypothetical protein BH1415 [imported] - Bacillus halodurans (strain C-125)
Db 29 EKCYWPLILET-EVPNMKVIAHNSGPLEWLNRPEYDILRSVLKRGVEIVVAGF 86	C;Species: Bacillus halodurans C;Accession: G33826 C;Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
Qy 79 THAILPLPLSRVHQVQDREVKE-ELFELSPKGFWNPELAYDPIPAILKDNGEYL 136	R;Takami, H.; Nakase, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Fuji, F.; H Nucleic Acids Res. 28, 4317-4331, 2000 A;Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a A;Reference number: A83650; MUID:20512582; PMID:11055132
Db 87 YEPVLAISIP---KEDIEQIRLMKWKASIGFARGWILTERWQPELYKTIKESGIDYV 143	A;Accession: G33826 A;Status: Preliminary A;Molecule type: DNA A;Residues: 1-943 <STO>
Qy 137 PADGEAMLFSAHLNSRIKPKPLY-PHLKAQRE----KRFY-TSYLGLRLERKA 187	A;Cross-references: GB:AP001512; GB:BA000004; NID:g10174030; PIDN:BA05134.1; GSPDB:G
Db 144 IVD------DVFHMAGLSKSEELVWVPPYTYEDGGEVIAVPPDEKLRL-YLIPFRPDVK 194	A;Experimental source: strain C-125 C;Genetics:
Qy 188 IKL---VFEGRVTLKAV-EDIEPAVPPVWAVNTAVMLGIRPLMPNPKVAVT- 236	A;Gene: BH1415
Db 195 LEYLHSLJDGEDESKVAVFHDDGEKGRGIWPCTYEWVY-----EKGWLREFDR 241	
Qy 237 ---EDKDNNILYGTIDIE----FIGYRDIAGYRMSEGVEGLIEVIDELNSELCLP----- 281	
Db 242 ISSDENLMLTEYERKVRGLTVFASVY-----FEM----FENSLPAKQARLF 289	
Qy 282 ---SELKHSG---RELYRITSSWAPDKSLSLRIWREDEGN---ARLNMLSYNMRGELL 330	Query Match Score 128; DB 2; Length 923; Best Local Similarity 20.1%; Pred. No. 0.16; Gaps 24; Matches 97; Conservative 60; Mismatches 167; Indels 158; Gaps 24;
Db 290 VEFVNEILKVGCFEKPRVFGIWN---KHF-FVYKYPESNYMHKRMVLNSKLVRNN--- 341	
Qy 331 AFNSDARGWELPERRLDAFRAIYND-WNG 359	Qy 3 ALVFHGNLQYAEIIPSE----IPKVIEKAYIPVIELIKIPFGLN-----46
Db 342 -----PEARYLLRAQCNDAYWIG 3660	Db 7 SLVLH AHLPHV RHQBEDR LPERW DEAMSETY PLLWALEKLVKHAYTISFTPPYMEML 66
RESULT 10 D71011	Qy 47 -----TGY-----TLKFLPKD-----IIDLVKG-----G 65
hypothetical protein PH1386 - Pyrococcus horikoshii	Db 67 SDPLYOTRYLNHLNTEQLKKERKTRNQRTNLQVQFQKQRTEKLATEFLDWRNLLIG 126
C;Species: Pyrococcus horikoshii C;Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000 C;Accession: D71011	Qy 66 TASDLE-----IIGTSYTHAILPLPLSR-VEAQVQDRREWYKELFELSPKGFWLPELA 119
R;Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Sekine M.; Ohnuki, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi DNA Res. 5, 55-76, 1998	Db 127 FRS-LMNEOQTLMSSAATTAFFPVLKTKREAIRAQVRGIACEQHFGKKPKGFWLPECA 185
A;Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeon: Pyrococcus horikoshii. A;Reference number: A11000; MUID:98344137	Qy 120 YDPPIIAILKDNGYEXLFADGEAMLF--SAHLNSAIRPIKPY-PH--LIKAQREKRF 172
A;Status: Preliminary; nucleic acid sequence not shown; translation not shown	Db 186 FSPGVDRILFEEGIRTYFVEDAHVILTADPUPHKGSS---APIYSPSGTIALFPRHTBLSA 241
A;Molecule type: DNA A;Residues: 1-560 <KAV> A;Cross-references: AP000006; NID:93236133; PID:BA030492.1; PID:93257809	Qy 173 RYTSYLLGLRERLRAKILYPEGKVTLKA-VKDEAVPVW-----VANTAVMLG 220
A;Experimental source: strain OT3 A;Note: this accession replaces an interim accession for a sequence replaced by GenBank C;Genetics:	Db 242 KWSSLIG-----YPGDVDFREFYRDAYDRENWDYKPHVKDGIRDTG-LK 288
C;Superfamily: Pyrococcus horikoshii hypothetical protein PH1386	Qy 221 IGRPLMNPKK---VASWIEDK-----DNILLYGTDIEFI 252
	Db 289 YRITHGTEKDLVYPREWAKVQEHANHFIGATHEDQHGQNFPYVWMPFDALF 348
	Qy 253 GYRDIAGRMSVEGLEV-----TDELNSELCPSELKSGRELYLTSSWAPDSKRIWR 308
	Db 349 GHWWFEGPEW-IEALYEQGADRVSITPELYLQRYQ-DFQTHVSFSTWGRDGYGHWVL 406
Query Match Score 130.5; DB 2; Length 560;	Qy 309 EDEGNARLNMLSYNMRGELL-----AENSNDARWEPLPERRLDAFRAIYN 355
Best Local Similarity 25.6%; Pred. No. 0.049; Mismatches 28; Conservative 60; Gaps 12;	Db 407 NDH-NAWMYRHMRMEKDLAKIVAMYQPQTVLKEQIAQMVREW-----MLAVSS 455
Db 71 EYKREFEKYMERKLKMEEDLERFDEKREAINF---MIGYF---KDVYSWIKID 122	RESULT 12 A71207 amyloplullanase PAB0122 - Pyrococcus abyssi (strain Orsay) C;Species: Pyrococcus abyssi C;Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Aug-1999 C;Accession: A75207 R;Anonymous Genoscope Submitted to the EMBL Data Library, July 1999 A;Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosomes S
Qy 14 EIKSEPKVIEKAYIVIEIL-----IKEEIPGLNITGYIILKFLKDIDL--VK 63	
Db 71 EYKREFEKYMERKLKMEEDLERFDEKREAINF---MIGYF---KDVYSWIKID 122	
Qy 64 GGIASDL-----IIGTSYTHAILPLPLSR-VEAQVQDRDEVKEBLFELSPKGFW 114	
Db 123 GNIGKFKRELQDEGGYEVITSRATHYPLLIGRDEAIEALLNGKVVYKYFGRKPRGFW 182	
Qy 115 LPETLAYDPI-----IPAILKDNGYEXLFADGEAMLFSAHNSATKTPK 158	
Db 183 LPECAYRDLWIKWPSPTGEVKWKGIEHEFLKFGIEFYFVE-----SHLIDK-GPYSL 234	
Qy 159 LYPLIKQOREKFRYTSYLGRLERKAIIVLVECKVTLKAVKQDIEATPVWVA 212	

A; Reference number: A755001

A; Accession: A75207

A; Status: preliminary

A; Molecule type: DNA

A; Residues: 1-1362 &lt;KAW&gt;

A; Cross-references: GB:AU248283; GB:AL096836; NID:95457433; PIDN:CAB49104\_1; PID:e151499

A; Experimental source: strain Orsay

A; Genetics: apu: PAB0122

Query Match 6.6%; Score 124; DB 2; Length 1362;

Best Local Similarity 21.5%; Pred. No. 0 55; Matches 82; Conservative 57; Indels 100; Gaps 16;

Query 32 SDLIEIIGTSYTHAILPL-----REKFRYISYLLGLRERLKAIKLVFEGKVTLKAVKDIIEAVPVW 21.0

Db 236 VETVILKQHWMLLNHT----PEEHKKNLLG--NGNEVTVYPXTHPIGPILNDFGW 286

Query 98 LSRVEAQYQDREYKEELF--ELSPKGFWLPLPELAYDPIIAIKLKDNGYEYLFDG--E 141

Db 287 YEDFDAQVKKANEELKEYLGAKTPKGWAADSLNDKTLEIANGWKWMIDDLVLE 346

Query 142 AMLFEAHLSAIKPK-----IKPLYPHLIKA 166

Db 347 KLGVPKTIESTYKPWVAQGDKKTYLPRNHDLSDRYGFRYAGMNOYDAVNFEELKI 406

Query 167 QREKPERVSYLIGSRELERKAIKVFEKGKVTLKAV-KDIEATPQVYAVNTAV----ML 219

Db 407 QRQNQDGSLVYVITDGENWEHWHPDGFGLFLDELYRQEEIQLQKGIRTYPSXEYLMF 466

Query 220 G-TIGRLPLMPKQKVAWSLEDKMLLGYLDIEFIGYDIAGYRMSEGVLEVIDELNSE 277

Db 467 GDKANLTPKMKRDEFTEEDNHALLKATL-----GLEYDKVGTB----- 509

Query 278 LCPLSELKHSQRELYLRTSSWAPPDKSLRIRWEDDEGNARLNMSYNMRGELLAENS-A 336

Db 510 -----EMQWPESSWI-IGTGSTW---IGEPQENIAWWLYLARKALEFKDNV 553

Query 337 RGWEWPRLERDRAFTAYNW 357

Db 554 DKWNKAYEY--LERAEGSDW 571

RESULT 13

AC1970 hypothetical protein air1310 [Imported] - Anabaena sp. (strain PCC 7120)

C; Species: Anabaena sp.

A; Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120

C; Accession: AC1970

R; Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuchi, S.

Nakazaki, N.; Shimapo, M.; Sugiimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.

DNA Res. 8, 205-213, 2001

A; Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana-

A; Reference number: AB1807; MUID:21595285; PMID:11759840

A; Accession: AC1970

A; Status: preliminary

A; Molecule type: DNA

A; Residues: 1-744 &lt;KUR&gt;

A; Cross-references: GB:BA000019; PIDN:BAB73267\_1; PID:g17130657; GSPDB:GN00179

A; Experimental source: strain PCC 7120

C; Genetics:

A; Gene: air1310

Query Match 6.6%; Score 123.5; DB 2; Length 744;

Best Local Similarity 22.6%; Pred. No. 0 25; Matches 79; Conservative 49; Indels 111; Gaps 19;

Query 68 SDLIEIIGTSYTHAILPL-----LPLSR-----VEAQYQDRKEELF 106

Db 208 SEQLEVTTPYTHPILPLADTNSGRAVNMALPESRFQWSEDPHRJLRAWELEYTERF 267

Qy	107 ELSPKGEFWLPLPELAYD-TIPAILKDNGYEYLFD-----GMAMLFSAHLSAIKPKIP--L 159
Db	268 GQDPKGWLWPSSESVSPDILPYIK-QGFQWICSDDEAVLGTLKHFHRGAGNYQQPELL 326
Qy	160 Y-PHLIKAQ-----REKFRYISYLLGLRERLKAIKLVFEGKVTLKAVKDIIEAVPVW 21.0
Db	327 YRPYRATPAGDIAITFRDHR--LSDLIGF-----TYGAMPQAAADL 368
Qy	211 VAVNTAVMLGIGNLPLNPKKASWEDKDNTLGYCTD---IEFTIGYDAGY-RMSYE 265
Db	369 VGHLOQAIXRMQRERPSEPWLPVIALDGENEWFYQDGKPFLEAL-YQSLSNEPHIKLV 427
Qy	266 GELLEVIDELNSCLPSLKHSGRELYLRTSSWAPDKSLRIWEDDEGNARLNMSYNMRG 325
Db	428 TVSFIEBEPATTIPAQLHQG-----SWV-LOGSFITWIGD----- 454
Qy	326 ELALLAENSNDARGMWEPLPERU-----DAFRAY-----NDW 357
Db	465 -----AKN--RAWDYLTEARMLNHPEATEBNPDEAWEALYAAEGSDW 506
RESULT 14	
Qy	B669553 methanol dehydrogenase regulatory protein (moXR) homolog - Archaeoglobus fulgidus
C; Species: Archaeoglobus fulgidus	
C; Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999	
C; Accession: B669553	
R; Klein, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Dod	
; Fleischmann, R.D.; Quackenbush, J.F.; Sutton, G.S.; Gill, S.; Kirnness, E.	
Gloede, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weiman, J.F.; McDonald, L.	
Nature 390, 364-370, 1987	
A; Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, Smith, H.C.; Woese, C.R.; Venter, J.C.	
A; Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing arch	
A; Reference number: A69250; MUID:98049343	
A; Accession: B669553	
A; Status: preliminary; nucleic acid sequence not shown; translation not shown	
A; Molecule type: DNA	
A; Residues: 1-124 <KLE>	
A; Cross-references: GB:AB001108; PID:6689431; PIDN:AB91247_1; PID:9265	
C; Superfamily: methanol dehydrogenase regulatory protein	
Query Match 6.4%; Score 120.5; DB 1; Length 324;	
Best Local Similarity 21.7%; Pred. No. 0 13; Matches 71;保守性 60; Mismatches 121; Indels 75; Gaps 14;	
Matches 71;保守性 60; Mismatches 121; Indels 75; Gaps 14;	
Qy	1 LRALYFGHNQYAYAIP--KSEIPKVIKEKAYIPVIELIKEIPFGUNITYTKEFLPKD 57
Db	35 LAAATGNTNFNLFVPLGKLTAKVFRV-----IGADYRRVQFTPDILPSD 82
Qy	58 IDLYKGGI-----ASDLIELI-GTSYTHAIL-----PLIPLSRVEAQYQDRKEELF 105
Db	83 II----GVKWRGDRFREVKGIFIFTNVYLADEINRSPPKTAALEMEKQIVGEET 137
Qy	106 FELSPKGFWLPLPELAYDPTPAIKLKDNGYEYLADGEAMLFSAHLSAIKPKIPLYHLIK 165
Db	138 FSLSMPFFVL-ATQNPI-----ELEGTPPLPEQMDRFLMRGPPIPESIE 182
Qy	166 AQREKRFRYISY-----LIGLRLRKAIKLVFGSKVTLKAVKDIIEAVPVW 211
Db	183 EMEELIRRISWRKDDPTEDVEPVYSLERTRIQDAEAVYDKSILKYYISELVR-----T 239
Qy	212 AVNTAVMLGI---GRPLMNPKVQASIEDKNTL--LYGTDIEFYGDRDAGYMSVE 265
Db	240 REHVELGGSPPGGLAKLKLARALAVMDGRDFVIPDVKRAVEALAHRYLKEFYAVE 299
Qy	266 GIL--EVIDELNSELCPLSPKLSKGRE 290
Db	300 GIRABEEVEELNSRVPR--KYBAQE 323

RESULT	15	Pyrococcus abyssi (strain Orsay)					
		:Species:	Pyrococcus abyssi				
		:Date:	20-Aug-1999	#sequence_revision	20-Aug-1999	text_change	20-Jun-2000
		:Accession:	C75120				
		:anonymous; Genoscope;					
		:submitted to the EMBL Data Library, July 1999					
		:Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome stru					
		:Reference number: A75001					
		:Accession: C75120					
		:Status: preliminary					
		:Molecule type: DNA					
		:Residues: 1-102					
		:Cross-references: GB:AT248285; GB:AL096836; NID:95458067; PID:CAB49676.1; PID:9545818					
		:Experimental source: strain Orsay					
		:Genetics:					
		:Gene: PAB1857					
		:Superfamily: Pyrococcus horikoshii hypothetical protein PH1386					
Query	6	Match	6	3%	Score	119;	DB 2;
		Best Local Similarity	26	8%	Pred. No.	0.41;	Length 602;
		Matches	42;	Conservative	20;	Mismatches	59;
						Indels	36;
						Gaps	4;
DY	14	EIPKEIPIKVEKAYIPVETLIKEEIPGLNITGYTLKFLPK-----					DII 59
Db	71	EYIKAFENRMRKTI----KLMRADGKIRNAATIIMFYFEDVYEWSKRINGDI					I 125
DY	60	DLVKGGIASDLIEIGTSYTHAILPLPLPSR-VEAQVQDFREYKEELPELSPKGFWNLPEL					118
Db	126	GRFKQIQLQDEGEVIIITSAAATHGYIPLLGDEAIQAQILTGIRVYEKYFGKKPRGIVWLPEC					185
DY	119	AYDPI-----IPAIKDKDGYEYLFDAD					139
Db	186	AYRPDLWKLWPSGTGEIKWKRGTEHPLKKLGLEFFEV					222

Search completed: June 2, 2002, 18:06:42  
Total time: 179 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2002, 18:05:58 ; Search time 29.22 Seconds

(without alignments), 2155.036 Million cell updates/sec

Title: US-09-886-400-4

Perfect score: 1877  
Sequence: 1 LRALYFHNLYQAEIPIKSEI.....RRLDAAFFAIYNDWRGENGEP 364

Scoring table: BLOSUM62

GapOp 10.0 , Gapext 0.5

Searched: 562222 seqs, 17294929 residues

Total number of hits satisfying chosen parameters: 5622222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTRIMBLE\_19:<\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_micr:\*

8: sp\_organelle:\*

9: sp\_phage:\*

10: sp\_plant:\*

11: sp\_rabbit:\*

12: sp\_rabbit:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_ivirus:\*

16: sp\_bacteriap:\*

17: sp\_archeap:\*

### RESULT 1

Q9HBB5 ID

PRELIMINARY; PRT;

364 AA.

AC Q9HBB5; PRELIMINARY;

DT 01-MAR-2001 (TREMBUREL, 16, Created)

DT 01-MAR-2001 (TREMBUREL, 16, Last sequence update)

DT 01-DEC-2001 (TREMBUREL, 19, Last annotation update)

DE ALPHA-GALACTOSIDASE.

GN GALA.

OS Pyrococcus furiosus.

OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.

NCBI\_TaxID=2261;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR EMBL; AF195244; AAC28455; 1; -

DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

### SEQUENCE

3E8E0AF5BDCCF2A5

CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_57.

DR Pfam; PF03065; Glyco\_hydro\_57; 1.

DR 3E8E0AF5BDCCF2A5 CRC64;

QP Q9HBB5 ID

SEQUENCE FROM N.A.

RC STRAIN=DSM 3638;

RA Verbose C.H.;

RL Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases.

DR DR InterPro; IPR004300; Glyco\_hydro\_





RN	[1] SEQUENCE FROM N.A.	KW	Hypothetical protein; Complete proteome.	
RP	SEQUENCE FROM JCM 10545 / 7;	SQ	SEQUENCE 895 AA; 101064 MW; 6BC7CD8380DDFBEC CRC64;	
RC	PubMed:11577479;	Query Match	Score 136.5; DB 17; Length 895;	
RX	Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S., Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y., Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Yamagishi A., Aoki K.-I., Masuda S., Yanagi M., Nishimura M., Oguchi A., Oshima T., Kikuchi H.; "Complete genome sequence of an aerobic thermoacidophilic Crenarchaeon, Sulfolobus tokodaii strain7.";	Best Local Similarity	7.3%; Pred. No. 0.02; Indels 95; Gaps 20;	
RA	"Complete genome sequence of an aerobic thermoacidophilic Crenarchaeon, Sulfolobus tokodaii strain7.";	Matches	85; Conservative 71; Mismatches 159;	
RA	DNA Res. 8:123-140 (2001).	Qy	3 ALVFGNLYQAEIPKSEPKVIEKAYIPVI----ETLIKE----IPFGINNT-----47	
RA	EMBL: AP000983; BAB65830.1; -	Db	377 SLVGAYELQALLQEVNNTID-FTVLLCYWETTILHEENATTSNGVNHSHDIEAV 434	
RT	DNA Res. 8:123-140 (2001).	Qy	48 GYTAKFLPKDIIIDLVKGIGASDLIELIGTSYHAILPIL---PLSRVEAQYORDREVKE 103	
RL	EMBL: AP000983; BAB65830.1; -	Db	435 NYTL-----NYLKT-LKEGVKEVLTVFPLQPLQDQGWSDYLQIEMGENMTH 486	
KW	Hypothetical protein; Complete proteome.	Qy	104 ELEELSPKGFWLPELAYDPLIAALKDNGYYEFADGBAML-FSAHNSAATPKIPLYPH 162	
SQ	SEQUENCE 443 AA; 53218 MW;	Db	467 EVFGTWANGTWTPEAFAFDMDLVGNECSISFILDQQAFLPVTYLNGSLNDQPFIVE 546	
Query Match	7.6%; Score 142.5; DB 17; Length 443;	Qy	163 LKQAREKRFRYISYSLGLRELRAKIVLVEGGKVTL--AVRDIEAVPVWWAVNTAVML 219	
Best Local Similarity	23.0%; Pred. No. 0.004;	Db	547 -----NNLGQTIIIVLFR-NTLSNEGFPEFFQSPLQTA--QELIQ 584	
Matches	79; Conservative 41; Mismatches 134; Indels 89; Gaps 15;	Qy	22 KVKLEKAYIPVIETLKEEIPPGLNITGYTERKFL-----PKDIIIDLVKGJIA 67	
Qy	46 RYKNKCYPATRILLBEIENG-ESEGVDKFEFFLSGTIFIEQAEKWGDVLEFOQLSY 103	Db	220 GIGRLPLMNPKKVASWIEDKDNLILYG----TDIEFIGYRDAGYRMSVGEGLIEDEL 274	
Db	68 SDLIELTGTSYHTAIIPLLPRLSRYIAQYQDREYKEEPELSLPELAYDPIIPA 126	Qy	585 QIAEITYMNPNPGGVVTVALDGGENPLIFNPITGPDSLYAI-YQALESEYQGS----WLYQT 638	
Qy	104 THKVEFLSQTYIHSITSLWDLTENKEQVQMHKNLIRDYFGQEVTFTENTELLTPRIVK 163	Db	275 NSELCLPSEKRSKSGRELYRTISWAPDKSERIRV-----EDENARANMMSYNNR 324	
Db	127 IJKDNGYEYLFDAGEMLFSAHNSAIKPKIPLYHLIKAQREKRFRYTSYLLGRELRK 186	Qy	639 ASEAI--ATHKPTTSITNLNVNSW-DLNLYNNINGYIGKTETWQNTSLAREXLIAVT- 693	
Qy	164 EIEKLGFKVITEGK-----ESLLKGKSP-----NRVYRIRDTKSLILLRNRYL 207	Db	325 GELALLAENSARGWEPL-PERRLDAFRAIYN-----DURGENEPL 364	
Db	487 ATKLVFESKVTULKAVIDEAIPVWVAVNTAVMLGIGRLPIMNPKVAVSIEDEDKNLILYG 246	Qy	694 ---ALGDNTISPVLVPFNTTPNSTDLDITLWMLVLYAEGSDTWQFgPP 739	
Db	208 SDDAFR-----FSNPKW-----DQYPL-TASKFADNVWSSEGFI---241	RESULT	8	
Qy	247 TDIEFIGYRDAGYRMSVGEGLLEVDELNSCELSPELKHSG-----RELYLR-----294	O50094	PRELIMINARY;	
Db	242 -GLFVDYETFGHHHWPESTGILDFL-----RWLPRRELHGVERKLPREVVYKEYDIEV 294	ID	560 AA.	
Qy	295 ---TSSWA---PDKS---LRIWRDEGEARNLMSYNNRGE 326	AC	PRT;	
Db	295 IDSTSVSWADINKDESSNLGNIMOWAYDEMVRTEMELAKEAGGE 337	DT	560 AA.	
OC	NCB_1386	DT	01-JUN-1998 (TRIMBLE). 06, Created)	
OS	Pyrococcus horikoshii.	DT	01-JUN-1998 (TRIMBLE). 06, Last sequence update)	
OC	Archaea; Buryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.	DT	01-JUN-2001 (TRIMBLE). 17, Last annotation update)	
OX	NCBI_TaxID=53953;	DE	HYPOTHETICAL PROTEIN PH1386.	
OX	"Complete sequence and gene organization of the genome of a hyper-	PH1386		
OX	thermophilic archaeobacterium, Pyrococcus horikoshii OR3."	OS		
OX	Archaea; Buryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.	RT		
OX	NCBI_TaxID=53953;	RT		
RN	[1]	RNA	SEQUENCE FROM N.A.	
RP	SEQUENCE FROM N.A.	RC	SEQUENCE FROM N.A.	
AC	Q972NO	STRAIN=JCM 10545 / 7;	RC	SEQUENCE FROM N.A.
DT	Q972NO; PRELIMINARY;	DT	STRAIN=JCM 10545 / 7;	
DT	01-DEC-2001 (TRIMBLE). 19, Created)	DT	01-DEC-2001 (TRIMBLE). 19, Last sequence update)	
DT	01-DEC-2001 (TRIMBLE). 19, Last sequence update)	DT	01-DEC-2001 (TRIMBLE). 19, Last annotation update)	
DE	HYPOTHETICAL PROTEIN ST1102.	DE	HYPOTHETICAL PROTEIN ST1102.	
GN	Sulfolobus tokodaii.	GN	Sulfolobus tokodaii.	
OS	Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.	OS	Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.	
OC	NCBI_TaxID=111955;	OC	NCBI_TaxID=111955;	
RN	[1]	RT	"Complete sequence and gene organization of the genome of a hyper-	
RP	SEQUENCE FROM N.A.	RT	thermophilic archaeobacterium, Pyrococcus horikoshii OR3."	
RC	SEQUENCE FROM N.A.	RT	DNA Res. 5:55-76 (1998).	
RX	Strain=JCM 10545 / 7;	DR	EMBL; AP000006; PAA30492.1;	
RA	PubMed:11572479;	DR	InterPro; IPR030493; DUF200.	
RA	Kawarabayashi Y., Hino Y., Horikawa H., Jin-no K., Takahashi M., Sekine M., Baba S.-I., Ankai A., Kosugi H., Hosoyama A., Fukui S., Nagai Y., Nishijima K., Otsuka R., Nakazawa H., Takamiya M., Kato Y., Yoshizawa T., Tanaka T., Kudoh Y., Yamazaki J., Yamagishi A., Aoki K.-I., Masuda S., Yanagi M., Nishimura M., Oguchi A., Oshima T., Kikuchi H.; "Complete genome sequence of an aerobic thermoacidophilic Crenarchaeon, Sulfolobus tokodaii strain7.";	DR	Pfam; PF02651; DUF200.1.	
RA	DNA Res. 8:123-140 (2001).	KW	Hypothetical protein; Complete proteome.	
RL	EMBL; AP000983; BAB66135.1; -	SQ	SEQUENCE 560 AA; 65769 MW;	
DR		Query Match	7.0%; Score 130.5; DB 17; Length 560;	
		Best Local Similarity	25.6%; Pred. No. 0.032;	
		Matches	60; Conservative 28; Mismatches 81; Indels 65; Gaps 12;	

QY	14 EIPSEIPKVIKEKAYIPIVIELT-----IKEEIPGLNITGYTLKFLPKDIDL---VK	63	QY	221 IGRPLMNPKK--VASWIEDK-----DNILLYGGTDIEFI	252
Db	71 EYTKREPFYMEERKLKNEEDLERFKDEKEAIFI-----NIGYF----KDViSYWKSID	122	Db	289 YHRITGITEEKDLYVREWAEKRVQHANPDKSLRlWR	348
QY	64 GGIASDL-----IEILTGSYTHAILPLPLPSR-VEAQVORDREYKEELFELSPKGFW	114	QY	253 GYRDIAGRMSVGGLEV---IDELENSELCPSELKHSGRELYLRTSSWADPKSLRlWR	308
Db	123 GNTLGKFRELQDGEYEVITSATAHGYLPLGRDEIEAQLINGKIVYEXYFGRKPRGFW	182	Db	349 GHWFEGPEW-IEALYEQADRSFSITPELYLQRHYQ-DEQTAIVSFTSWGRIGYGHWL	406
QY	115 LPELAYDPI-----PAIKLNGYZAYLPLADGEAMLSAHLSAIKPIK 158		QY	309 EDEGANLNMLNSYNMRGELA-----AENDARGWEPLPERRLDAFRATYN	355
Db	183 LPECAYRDPDGWLKSPSTGEVKWRKGEEHFLKGFIYFFEFYF-----SHLIDK-GPVSL	234	Db	407 NDH-NAVMYRHMRDLAKTAVAMYQPVTYLEKAQIQMVREN-----MLAVSS	455
Qy	159 LYPHLIIKQREKFRFESYLLGLRELEKAIKVFEKVLKAVKDIEAIVPWVA	212	QY	356 DW 357	
Db	235 RYGNLPLA-KTKRSTLRFYF----LKNGIAVFARRRET-----GICQWWSA	274	Db	456 DW 457	
<hr/>					
RESULT 9					
Q9KDD4	PRELIMINARY;	PRT; 923 AA.	Q97294	PRELIMINARY;	PRT; 1352 AA.
ID	Q9KDD4		ID	Q97294	
AC	Q9KDD4;		AC	Q97294;	
DT	01-OCT-2000 (TREMBLe1, 15, Created)		DT	01-MAY-2000 (TREMBLe1, 13, Created)	
DT	01-OCT-2000 (TREMBLe1, 15, Last sequence update)		DT	01-MAY-2000 (TREMBLe1, 13, Last sequence update)	
DT	2001-06-01 (TREMBLe1, 17, Last annotation update)		DT	2001-06-01 (TREMBLe1, 19, Last annotation update)	
DE	BH1415 PROTEIN.		DE	ANTLOPUULUANASE.	
GN	BH1415.		APU OR PAB0122.		
OS	Bacillus halodurans.		Pyrococcus abyssi.		
CC	Bacteria; Firmicutes; Bacillales; Clostridium group;		Archaea; Euryarchaeota; Thermococcales; Pyrococcus.		
OC	Bacillus;Staphylococcus group; Bacillus.		OC	NCBI_TaxID=29292;	
OX	[1]		RN	[1]	
RN	NCBI_TaxID=86665;		RP	SEQUENCE FROM N_A.	
RP	FROM N_A.		RC	SEQUENCE FROM N_A.	
RC	STRAIN=C-125 / JCM 9153;		RA	Pyrococcus abyssi genome sequence: insights into archaeal chromosome	
RX	MEDLINE=2012582; PubMed=11058132;		RA	structure and evolution.",	
RA	Takami H., Nakasone K., Takeki Y., Maeno C., Sasaki R., Masui N.,		RA	Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.	
RA	Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kohara S.,		DR	EMBL; AU48283; CAB49104; InterPro; IPR0430; Pfam; PF3065; Glyco_hydro_57.	
RA	Horiuchi T.;		DR	DR	
RT	Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and genomic sequence comparison with Bacillus subtilis.",		RT	Complete proteome.	
RL	Nucleic Acids Res, 28:4317-4331(2000).		RN	SEQUENCE 1352 AA; 154578 MW; 277AFAB4E14860D1 CRC64;	
DR	AP011512; BAB0134; .		RC	SEQUENCE FROM N_A.	
DR	InterPro; IPR003803; DUF200.		RA	Pyrococcus abyssi genome sequence: insights into archaeal chromosome	
DR	InterPro; IPR001296; Glycos_transf_1.		RA	structure and evolution.",	
DR	Pfam; PF0534; Glycos_transf_1.		RA	Submitted (JUL-1999) to the EMBL/GenBank/DDBJ databases.	
DR	Pfam; PF0534; Glycos_transf_1.		DR	EMBL; AU48283; CAB49104; InterPro; IPR0430; Pfam; PF3065; Glyco_hydro_57.	
KW	Complete Proteome.		DR	DR	
SQ	SEQUENCE 923 AA; 106638 MW; 8C7BD64F3603553 CRC64;		RT	Complete proteome.	
<hr/>					
Query Match	6.8%	Score 128; DB 16; Length 923;	Query Match	6.6%	Score 124; DB 17; Length 1362;
Best Local Similarity	20.1%	Pred. No. 0.11; Indels 158; Gaps 24;	Best Local Similarity	21.5%	Pred. No. 0.39; Mismatches 142; Indels 100; Gaps 16;
Matches	97	Conservative	Matches	82	Conservative
Qy	3 ALVFGNLYQAYEIKSE----IPKVEKAYIPVETLIREIPGLNI-----46		Qy	32 IETIJKIEEPGGINITYLKLPKDIDLYKGIGIASLDLIEIGTSYTHAILPLP-----87	
Db	7 SLVLEAHLPYVRHQEDLLEERNLFEMSETYIPLWALEKLPLVKHAVTISFTPPYMEML 66		Db	236 VETVLIKQH-WLHNHT----FEHEKINLLIG---NGNVEKQVTPYTHIGPLINDFGW	286
Qy	47 -----TGY-----TLEKPLKD-----IDLVKG-----G 65		Qy	88 LSRYVEAQVORDKEKEELF---ELSPKGFWLPELAYDPLIPAIKDNGYEYLADG---E 141	
Db	67 SDPLQTQTRYLNLTNEOLQKKEERKTNDQNLVQFYKQREYKELKATELOWDRMLIG 126		Db	287 YEDDAQYKKANELEYKEYLGACKVTPKGWWAEASALNDKTLELAENGKWWMDQVL	346
Qy	66 IASDIE-----IIGTSYTHAILPLPLPSR-VEAQVORDREYKEELFELSPKGFWLPELA 119		Qy	142 AMIFSAHNSAIKP-----IKPLPHHLIK 166	
Db	127 FRS-LMENEQQTLMMSAATHAFFPLKTKEAIRAQVRHGIACEQFGKKPLGFWLPECA 185		Db	347 KIGVPKTIESYYRPWQAQGDKKIYLFRNHDLSDRVGRYAGMNOYDAYKNFYEELKI 406	
Qy	120 YDPIIPLAIIKDGYZYLFDGEAMLF--SAHLSAIKPIK-Y-PH--IJKAQREKRF 172		Qy	167 QREKFRYISYLGLRERLKAJMLYFGKVTAV-KDIEAVPVVAVNTAV-----ML 219	
Db	186 FSPGVDRILPEEGTYFDEHAVLTADPYPHKGS---APIYSGIAJPRHTELSA 241		Db	407 QKQNIDGSLVYVITLDGENPWEHYFPDKLFLELRQEELQKGLIRIVTPSEYIEMF 466	
Qy	173 RYISYLLGIRLKRNIKLYPEGKYLKA-VKDIAVPVW-----VAVNTAVMLG 220		Qy	220 G-TIGRLPLMNPKKVAVSHIEDKNLLYGTDFIYGRTDAGYRMSVEGLEVDELNSE	277
Db	242 KVNSSTLG-----YPGDVYDREFYRDAYDREWDT-KPHVHKGDIRIDIG-LK 288		Db	467 GDKANKLTKMMRDLDETTDNVNLALKRKT-----GELYDMVGTE-----509	
<hr/>					
Qy	221 IGRPLMNPKK--VASWIEDK-----DNILLYGGTDIEFI	252	Qy	278 LCPLSEKHSGRELYLRSSWADSKSRlWREDEGNARLNLSYNNRGELALLAENS-D-A	336
Db	289 YHRITGITEEKDLYVREWAEKRVQHANPDKSLRlWR	348	Db	510 --ENQWPESSIWI-DGTLSTW---IGEPQENTAWNLYLARKALFENKDNY 553	
Qy	337 RGWPBLPERRLDAPRATYN 357		Qy		



DR	EMBL; AJ248285; CAB49676.1; -.	Qy	65 GIASDL-----IEIIGTSYTHAILPL--IPLSRVEAQVORDREYKEELFELSPKGF
DR	InterPro; IP001395; Aldo_ket_red.	;	113
DR	InterPro; IP003303; DUF200.	DB	130 ALLERINHEFFRTGSIELLATAVNCFLPYQDMES-ISAQIXMGLNRYRKHSISIPRGF
DR	SMART; SM00278; HHHL_1.	Qy	114 WLPELAYDPIPAALKDNGEYLADGEAMLFSHL 149
DR	PROSITE; PS00062; ALDOKETO_REDuctase_2; UNKNOWN_1.	DB	189 YLPELYGAPALERTIKSYSFTSYTLETHSFLFGTRV 224
KW	Hypothetical protein; Complete proteome; Sequence 602 AA; 70219 MW; 326B198D54DD43E CRC64;		
SQ			
		RESULT 15	
		Q97ZD2 PRELIMINARY; PRT; 447 AA.	
		ID Q97ZD2	
		AC Q97ZD2	
		DT 01-OCT-2001 (TREMBrel. 18, Created)	
		DT 01-OCT-2001 (TREMBrel. 18, Last sequence update)	
		DT 01-OCT-2001 (TREMBrel. 18, Last annotation update)	
		DE ALPHA AMYLASE (EC 3.2.1.1).	
		SSO0988	
		OS Sulfolobus solfataricus.	
		OC Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.	
		OC NCBI_TaxID=2287;	
		RN [1]	
		RP SEQUENCE FROM N.A.	
		RC SPRAINATCC 35092 / DSM 1617 / P2;	
		RX MEDLINE-21322986; PubMed-11427726;	
		RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,	
		RA Awayez M.J., Chan-Weber C.C.Y., Clausen I.G., Curtiss B.A.,	
		RA De Moors A., Eraso G., Fletcher C., Gordon P.M.K.,	
		RA Heijkamp de Jong I., Jeffries A.C., Kozaera C.J.J., Medina N., Peng X.,	
		RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,	
		RA Charlebois R.L., Doolittle W.F., Duguet M., Gaasterland T.,	
		RA Garrett P.A., Regan M.A., Sensen C.W., Van der Oost J.,	
		RT The complete genome of the crenarchaeon Sulfolobus solfataricus P2. ";	
		RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840 (2001).	
		DR EBO06718; AAK41260.1; -	
		KW Hydrolase; Glycosidase; Complete Proteome.	
		SQ SEQUENCE 447 AA; 53641 MW; E3A45F31AC2D20EE CRC64;	
		Query Match Score 5.8%; Best Local Similarity 20.8%; Pred. No. 1.2; Mismatches 72; Conservative Matches 72; Length 447;	
		Query Match Score 5.8%; Best Local Similarity 20.8%; Pred. No. 1.2; Mismatches 135; Indels 85; Gaps 17;	
		Qy 17 KSBIPKVIKEAYTPVETLI-----KEE--IPFGLNITGYTFLP--KDIDLIVKG 64	
		Db 41 KEFPERIKKNCTPATNIISSIEREEGVNVKPFSTSGTQEAEWRGCTIEFLQQ 100	
		Qy 65 GIASDLIEIGTSYTHAILPL--PLSRVEAQVORDREYKEELFELSPKGFWLPLAYDPI 123	
		Db 101 LAWKHKBFLAQTYHHVVTSLWEDKSEWKEVKMERDTISYFGCOPTFENTL--- 155	
		Qy 124 IPMLKDNGYEYLADGEAMLFSAHNSA1KPKIPLYPHJKAREKRFRYISYLGLRE 183	
		Db 156 ---ITKDIVEEVKMFKMLSEGTMNRNDRGSPNYVILKG-HEIRALFRNTLS--- 208	
		Qy 184 LRAKAIALYFEGKVTLKAVKDI- AVPIWWAVNTAVMLGIGRPLMNPKKVASWIEDKD 240	
		Db 209 -----DDIAFRFPNPNW-----DQPL-TASKYAWISRE 238	
		Qy 241 NLLYGIDIEFIGYRDAGYRNSVEGLEVID-- EILNS--EILCLPSELKHSGREY- 292	
		Db 239 G----NYGLIFYETGEHHIRETGILEFLKWLPTELNSKGVMMPREYVN--DYD 291	
		Qy 293 ---LRTSSWA---PDKS---LRTIWREDEGARNLNMNSYNMGE 326	
		Db 292 EIEIAHTISWADIEDEKSWLGNMQWAYDAVRRAEMPSRELQNE 337	
		Search completed: June 2, 2002, 18:09:36	
		Job time: 218 sec	
DR	EMBL; AJ248285; CAB49676.1; -.	Qy	3 ALVFGNLOYAE-----IPKSBEIPKVIKEAYTPVETLI-----ETLIKEBIPGLNI----- 46
DR	InterPro; IP001215; AAC65344.1; -.	Db	10 AFVDCNLPFVRGAGASSLLAESREFLEISITYPLLRLCETLERRVPPNISLAIGPV1 69
DR	TIGR; TP0358; -.	Qy	47 -----TGYTFL--PKDIDLVKG 64
DR	InterPro; IPRO03803; DUF200.	Db	70 CEMLANRVLMDRYRALDALIEGEREAIRLRSLSQERVQEAFLRSLSRSHRDYFDHCDG 129



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - Protein search, using sw model

Run on: June 2, 2002, 18:01:42 ; Search time 32.53 Seconds

(without alignments)

1242.879 Million cell updates/sec

Title: US-09-886-400-4

Perfect score: 1877

Sequence: 1 LRALYFGNLYQAYEIPKSRIL...RRLDAAFRAYNDWRGENGEPE 364

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Genesseq\_052802,\*

1: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1980.DAT:\*

2: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1981.DAT:\*

3: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1982.DAT:\*

4: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1983.DAT:\*

5: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1984.DAT:\*

6: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1985.DAT:\*

7: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1986.DAT:\*

8: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1987.DAT:\*

9: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1988.DAT:\*

10: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1989.DAT:\*

11: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1990.DAT:\*

12: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1991.DAT:\*

13: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1992.DAT:\*

14: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1993.DAT:\*

15: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1994.DAT:\*

16: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1995.DAT:\*

17: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1996.DAT:\*

18: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1997.DAT:\*

19: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1998.DAT:\*

20: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA1999.DAT:\*

21: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA2000.DAT:\*

22: /SIDS1/geodata/hold-geneseq/geneseqp-emb1/AA2001.DAT:\*

## RESULT

1 AAW34643 standard; Protein: 364 AA.

AAW34643;

XX

AC

XX

DT

27-MAR-1998 (first entry)

Thermostable alpha-galactosidase AEDII12RA alpha-gal-18GC.  
XX  
Alpha-galactosidase; alpha-glycosidase; thermostable enzyme;  
DE food processing; alpha glycoside hydrolase; raffinose;  
KW stachyose; verbascone; bean; flatulence; AEDII12RA-alpha-gal-18GC.  
KW  
XX  
OS Thermococcus alcaliphilus strain AEDII12RA  
XX  
PN WO9732974-A1.

Key difference 329  
FT Misc-difference 329  
FT /note= "Encoded by CTT"

## SUMMARIES

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Match	Length	DB ID	Description
1	1873	99.8	364	18	AAW34643	Thermostable alpha-glycosidase; thermostable enzyme; alpha-galactosidase; alpha-glycosidase; thermostable enzyme; alpha-glycoside hydrolase; raffinose; stachyose; verbascone; bean; flatulence; AEDII12RA-alpha-gal-18GC.
2	144.5	7.8	655	22	AAB96088	Putative P. abyssi Pyrococcus furiosus Putative P. abyssi Super heat resistant S. epidermidis Ope S. epidermidis Ope Heat resistant mal Arabidopsis thaliana Arabidopsis thaliana
3	143.5	7.6	649	15	AA47304	
4	124.4	6.6	1162	22	AAB96591	
5	121.5	6.5	653	19	AAW54870	
6	104.5	5.6	360	22	AGB82536	
7	104.5	5.6	376	22	AGB82536	
8	104	5.5	329	17	ARR94013	
9	104	5.5	571	21	AGI16038	
10	104	5.5	571	21	AGA48655	
11	104	5.5	704	21	AGI16037	

N-PSDB, AAT93753 .

XX Nucleic acid encoding alpha-glycosidase from Thermococcus  
PT alcaliphilus - used in food processing to hydrolyse  
PT alpha-glycosides, e.g. raffinose  
XX Claim 1; Fig 1; 32pp; English.

CC This protein comprises AEDII12RA-alpha-gal-18GC, a claimed  
CC thermophilic alpha-galactosidase of Thermococcus alcaliphilus  
CC (see AEDII12RA). Also claimed are: (1) an isolated polynucleotide (see  
CC PH 9.5; AAI93753) encoding the alpha-galactosidase; (2) a vector containing  
CC the polynucleotide or homologous or complementary sequences; (2)  
CC host cells containing the vector; (3) a process for producing the  
CC alpha-galactosidase in transformed or transfected host cells; an  
CC enzyme showing at least 70% identity to alpha-galactosidase and  
CC comprising at least 30 amino acid residues of its sequence; and (4)  
CC a method for hydrolysing alpha-galactoside bonds using the enzyme.  
CC The enzyme can be used to hydrolyse raffinose to sucrose and glucose  
CC in sugar beet processing (raffinose inhibits crystallisation of  
CC sucrose), and as a digestive aid to hydrolyse raffinose, stachyose  
CC and verbascose in beans and other gassy foods.

XX Sequence 364 AA;

Query Match 99.8%; Score 1873; DB 18; Length 364;  
Best Local Similarity 99.7%; Pred. No. 1.8e-182;  
Matches 363; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 LRAVIFHNLQVAEIPKSEIPKVIKRAYIPVETLKEEIPFGUNITGTLKFLPKDID 60  
Db 1 lralvfgnlqlqraipkseipkvikeyipvietlkeeipfgunitgtlkflpkdid 60  
Qy 61 LYRGIASDLIEIGTSYTHAILPLPSRVEAQYORDREKEYELFELSPKGFWPELAY 120  
Db 61 lvrgiasdlieigtsythailplpsrveaqyordrevkeefelspkgfwpeay 120  
Qy 121 DPITPAIKKDNGYEYLFDAGEAMLSAHLNSAIKIPKPLYLIKACRERFRYTSYLG 180  
Db 121 dpitpaikkdngyeylefdageamlsahlnsaikipkplylikacrekfrfytsylg 180  
Qy 181 LREIRKAIIKLYFEGKVTKAYDKDIEAVPVWAVNFAVMGLGTRPLMPKVKASWIEDKD 240  
Db 181 lreirkaiklyfegkvtkaydkdiedavpvwavnfvamglgrplmpkvkaswiedkd 240  
Qy 241 NIJYGTIDIEFIGYRDAGYRMSVGSILEVIDELSELCLPSETKHSGRELYIERTSSWAP 300  
Db 241 nijygtidiefgyrdagymsvgsilevidelcelcplsetkhsgrelyiertsrap 300  
Qy 301 DKSRIWIREDGEARLMLSYNMGRGELALLAENDSARGWEPPLPERRLDAFRAYNDWRE 360  
Db 301 dksriwiredegarlmlsynmgrgelallaeasnargwpplperrldafrayndwre 360  
Qy 361 NGEP 364  
Db 361 ngep 364  
RESUL 2  
AAB96088 standard; Protein; 655 AA.  
XX Putative P. abyssi alpha-amylase.  
DE XX Hyperthermophilic archaeon; hyperthermophilic protein.  
KW XX Pyrococcus abyssi.  
OS XX

PN FR2792651-A1.  
XX 27-OCT-2000.  
PD XX 21-APR-1999; 99FR-0005034.  
PR XX 21-APR-1999; 99FR-0005034.  
PA (CNRS CENT NAT RECH SCI  
PA (IFREMER INST FR RECH EXPL MER.  
PI Forterre P, Thierry JC, Prieur D, Dietrich J, Leconte O;  
PI Querellou J, Weissenbach J, Saurin W, Heilig R;  
DR WPI: 2001-126236/14.  
XX New nucleotide sequences isolated from Pyrococcus abyssi encode  
PT proteins useful in industry -  
XX Claim 7; Pages 711-712; 1657PP; French.  
PS Sequence 655 AA;

Query Match 7.8%; Score 145.5; DB 22; Length 655;  
Best Local Similarity 22.4%; Pred. No. 1.1e-05;  
Matches 90; Conservative 55; Mismatches 127; Indels 129; Gaps 22;

Oy 25 EKAYIPVETLKEEIP---FGLNITGTYTLKFLPK---DIDIVYKGGIASDLIBIIGTSY 78  
Db 28 ekayipflel-eeypnmkvainsgivewneenkpkvaidlksivrkgrveivragf 85  
Oy 79 THAIPPLPLSRVEAQYORDREVKEEIPFGUNITGTLKFLKONGYEYLFA 138  
Db 86 yepvilaipkedrcleyllkewakki-eydaksgwltervwpelktireagievvv 144  
XX SQ Sequence 655 AA;

Query Match 7.8%; Score 145.5; DB 22; Length 655;  
Best Local Similarity 22.4%; Pred. No. 1.1e-05;  
Matches 90; Conservative 55; Mismatches 127; Indels 129; Gaps 22;

Oy 139 DCEAMLESAHNSAIKPKPLYLHKQRE-----KFRY-----TSYLL 179  
Db 145 d-----dyfmagslkdqifwpytedgevitvpidekrylipfrprkviyah 198  
Oy 180 GIRELRKAIIKLYFEGKVTKAYDKDIEAVPVWAVNFAVMGLGTRPLMPKVKASWIEDKD 239  
Db 199 siasedeskavfh-----dddekgfjw-----pm-----tyerryek 231  
Oy 240 D-----NILLYGTDIE----FIGYRDAGYRMSVGSILEVIDELSELCLPSETKHSGRELYIERTSSWAP 323  
Db 232 gwleffgrssdeainimlysey.lqfkfkpglypiasyfemsews1paqqaklive- 290  
Oy 267 LLEVIDRLNSELCLPSELKHSGRELYLRTSSWAPDKSIRIWEDEGN--ARNMMLSYNM 323  
Db 291 fexkeln----mferyrfvrgiw---knf fykypeanyrhkrmlms--- 334  
Oy 324 RGELALIAENSDARGWEPPLPERRLDAFRAYND--WRGENG 362  
Db 335 ---rlrdnparrf -----viragondawhngvfg 362

RESULT 3  
AAR47504 standard; protein; 649 AA.  
ID AAR47504  
ID AAR47504 standard; protein; 649 AA.  
XX AC AAR47504;

XX XX XX AC AAB96091;  
 DT XX DT 29-OCT-2001 (first entry)  
 DE XX  
 Pyrococcus furiosus alpha amylase.  
 KW XX Putative P. abyssi amylopolullanase.  
 KW XX Hyperthermophilic archaeon; hyperthermophilic protein.  
 OS XX OS Pyrococcus abyssi.  
 XX PN FR2792651-A1.  
 XX PD 27-OCT-2000.  
 XX PF 21-APR-1999; 99FR-0005034.  
 XX PR 21-APR-1999; 99FR-0005034.  
 XX PA (CNRS ) CNRS CENT NAT RECH SCI.  
 PA (IFRE-) IFREMER INST FR RECH EXPL MER.  
 XX PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;  
 PI Querellou J, Weissbach J, Saurin W, Heilig R;  
 XX WPI; 2001-126236/14.  
 XX PT New nucleotide sequences isolated from Pyrococcus abyssi encode  
 PT proteins useful in industry -  
 XX PS Claim 7; Pages 715-719; 1657pp; French.  
 XX CC The present invention relates to the genomic sequence of Pyrococcus  
 CC abyssi (see AAF86131 and AAH41243-7) and P. abyssi proteins. P. abyssi is  
 CC a hyperthermophilic archaeon, which is isolated from deep-sea  
 CC hydrothermal vents. The present sequence is one such P. abyssi protein.  
 CC The proteins of the present invention have various potential industrial  
 CC uses, since the proteins are stable at very high temperatures, some up to  
 CC 110 degrees centigrade.  
 CC Note: This patent is in the same patent family as WO200065062, which  
 CC contains additional sequences as shown in AAB9132-AAH9143,  
 CC AAH7903-AAH75920 and AAG66136.  
 XX SQ Sequence 1362 AA;  
 Query Match Score 143.5; DB 15; Length 649;  
 Best Local Similarity 22.8%; Pred. No. 1.8e-05;  
 Matches 91; Conservative 59; Mismatches 118; Indels 131; Gaps 24;  
 QY 25 EKAYIPVIETLKEEPIP---FGLNITGYTLKFLPKDI---IDLVKGGIASDLIEIGTS  
 26 29 ekcypflel--eeypnkvalnqspiewi-qdrpeydlrlsvkrqgvevvag 85  
 DB 78 YTHAIIPLPLLSRVEAQVORDREKE--ELFELSPKGWLPELAYDPLIPALKNDGYEX 135  
 DB 86 fyepvlasip--kedriegirlnkewaksigfargwlttervwpeivktilegidy 142  
 QY 136 LFADGAMLSAHLNS-----AIKP KIRPLPHIKAQREKRFRYSYLL 179  
 DB 143 vvd-dyhfmseaiskeelywyppyteddgevavfpid-----ek---iryli 186  
 QY 180 GLRELRKAIKL---VFEKGVTLKAV--KDIAEVPVWVAVNTAVMLGGRPLIMNPKVVAS 234  
 DB 187 pfrpvckvlehlslidgdsksavfhddgekfgiwptwyewy-----ekg 233  
 QY 235 WI-----EDKDNILLYGTDE----FIGYDIAGRYMSVEGLLEVIDELNSBLCL 280  
 DB 234 wifrefdrissdeknimlyteleykyprgivylasay-----fem----sews1 281  
 QY 281 P-----SEIKHSG ---RELYIIRTSSWAPDKSLRIWREDEGN---ARLNMLSYN 322  
 DB 282 pakqarlfvefvne\_kvkgifekyrvfvrqgiw--knf-fytypesymhkrmimvsk1 337  
 QY 323 MRGEALLAENSDARGWEPULPERRLDAFFRAYND-WRG 359  
 DB 338 vrnn-----peazkylliraqcndayng 360  
 QY RESULT 4  
 ID AAB96091 standard; Protein; 1362 AA.  
 DB 510 -----emqwpesswi-dgt1stw--igesgeniawyiy-larkalfenkdnv 553

QY 337 RGWEPLPERRLDAFRAYNDW 357  
 :||| :||| :||| :||| :|||  
 Db 554 kwnkayey---lraegsdw 571 .

RESULT 5  
 AAW4870 ID: AAV54870 standard; protein: 653 AA.

XX AC AAG82528;  
 XX DT 01-SEP-1998 (first entry)  
 XX DE Super heat resistant 4-alpha-glucanotransferase.  
 XX KW Super heat-resistant 4-alpha-glucanotransferase; heat-treatment;  
 XX alpha-1,4-glucan; alpha-1,4-glucoside bond.  
 XX OS Pyrococcus sp.  
 XX PN JPI150986-A.  
 XX PD 09-JUN-1998.  
 XX PF 21-NOV-1996; 96JP-0311117.  
 XX PR 21-NOV-1996; 96TP-0311117.  
 XX (BEAB-) BE ABLE KK.  
 PA (NAGA-) NAGASE SEIKAGAKU KOGYO KK.  
 XX DR WPI: 1998-373989/33.  
 DR N-PDB: AAV2026.  
 XX PT New 4-alpha-Glucanotransferase which has been heat-treated - used at  
 PT high temperatures to transfer at least one glucose unit  
 XX PS Claim 4; Page 10-13; 18pp; Japanese.  
 XX CC The super heat-resistant 4-alpha-glucanotransferase has an optimum pH of  
 CC 6.0-8.0, with an optimum temperature at pH 7.5 of 100 deg. C. It has at  
 CC least 90% activity after heat-treatment at 100 deg. C for 30 minutes at  
 CC pH 7.5. It can transfer at least one glucose unit in alpha-1,4-glycan  
 CC to alpha-1,4-glucoside bond.  
 XX SQ Sequence 653 AA;

Qy 282 SELKHSGRELYRTRSSWADKSSRIWREDEGN--ARLNMLGYNMRGELALLAENS DARG 338  
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
 Db 297 tenkfdryfrvragiw---knf fffky pesymhkmvlmvskavron----- 340  
 Qy 339 WBLPLPERRLDAFRAYND-WRGENG 362  
 :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||  
 Db 341 ---peareflraqcndaywhgrfg 362

RESULT 6  
 AAG82528  
 ID AAG82528 standard; protein: 360 AA,  
 XX AC AAG82528;  
 XX DT 03-SEP-2001 (first entry)  
 XX S. epidermidis open reading frame protein sequence SEQ ID NO:2150.  
 DE XX staphylococcus epidermidis SRI strain; infection; diagnosis;  
 KW vaccination; endocarditis  
 KW staphylococcus epidermidis.  
 OS XX  
 PN WO200114809-A2.  
 XX PD 17-MAY-2001.  
 XX PR 09-NOV-2000; 2000WO-US30782.  
 XX PR 09-NOV-1999; 99US-0164258.  
 XX PA (GLAX ) GLAXO GROUP LTD.  
 XX PI Kimmerly WJ;  
 XX DR WPI: 2001-316495/33.  
 XX N-PSDB; AAH33378.

PT Nucleic acids encoding polypeptides from *Staphylococcus epidermidis*,  
 PT useful for vaccinating against infections, e.g. endocarditis -  
 XX Claim 18; Page 578-579; 2188pp; English.

CC AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides  
 CC (II), given in AAG81454 to AAG83120, from *Staphylococcus epidermidis*.  
 CC (I) and (II) can have antibacterial activity and therefore can be used  
 CC in vaccination. The nucleic acids (I) may be used to produce the  
 CC *S. epidermidis* polypeptides (II) via the production of vectors  
 CC containing them which are used to produce host cells which express the  
 CC polypeptides. The polypeptides (II) (and/or nucleic acids) may then be  
 CC used to vaccinate subjects and to raise antibodies against the bacteria.  
 CC The polypeptides may also be used to assay for other inhibitors of their  
 CC activity and therefore identity compounds that may be used for the  
 CC treatment of *S. epidermidis* infections, e.g. endocarditis. AAH53971 to  
 CC AAH55090 represent specifically claimed *S. epidermidis* genomic DNA  
 CC polynucleotide sequences from the present invention. AAH55091 to  
 CC AAH55098 represent oligonucleotide sequences and primers which are used  
 CC in the exemplification of the present invention.  
 CC N.B. The present invention specifically claims all the polynucleotide  
 CC sequences given in the sequence listing of the present specification,  
 CC however the sequence listing only goes up to SEQ ID NO:4454 so even  
 CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,  
 CC no sequences are present for SEQ ID NO:4455 to 4464.  
 XX SQ Sequence 360 AA;

Query Match Score 104.5; DB 22; Length 360;  
 Best Local Similarity 19.4%; Pred. No. 0.069;  
 Matches 58; Conservative 51; Mismatches 117; Indels 73; Gaps 12;

Qy 4 LVFHGNLQYAEIPKSEIPKVIKPVIEKAYIPVIEIIPFGUNITGYTLKFLPKDIDLVK 63





27-JUL-1999; 990US-0145913.  
27-JUL-1999; 990US-0145918.  
28-JUL-1999; 990US-0145951.  
02-AUG-1999; 990US-0146386.  
02-AUG-1999; 990US-0146388.  
02-AUG-1999; 990US-0147038.  
04-AUG-1999; 990US-0147204.  
04-AUG-1999; 990US-0147302.  
05-AUG-1999; 990US-0147192.  
05-AUG-1999; 990US-0147260.  
06-AUG-1999; 990US-0147303.  
06-AUG-1999; 990US-0147419.  
09-AUG-1999; 990US-0147493.  
09-AUG-1999; 990US-0147935.  
10-AUG-1999; 990US-0148171.  
11-AUG-1999; 990US-0148319.  
12-AUG-1999; 990US-0148341.  
13-AUG-1999; 990US-0148565.  
16-AUG-1999; 990US-0148684.  
17-AUG-1999; 990US-0149368.  
17-AUG-1999; 990US-0149175.  
18-AUG-1999; 990US-0149426.  
20-AUG-1999; 990US-0149723.  
20-AUG-1999; 990US-0149929.  
23-AUG-1999; 990US-0149902.  
23-AUG-1999; 990US-0149930.  
25-AUG-1999; 990US-0150566.  
26-AUG-1999; 990US-0150884.  
27-AUG-1999; 990US-0151065.  
27-AUG-1999; 990US-0151066.  
27-AUG-1999; 990US-0151080.  
30-AUG-1999; 990US-0151303.  
31-AUG-1999; 990US-0151438.  
01-SEP-1999; 990US-0151930.  
01-SEP-1999; 990US-0152363.  
10-SEP-1999; 990US-0153070.  
13-SEP-1999; 990US-0153758.  
15-SEP-1999; 990US-0154018.  
16-SEP-1999; 990US-0154039.  
20-SEP-1999; 990US-0147779.  
22-SEP-1999; 990US-0155139.  
23-SEP-1999; 990US-0155486.  
24-SEP-1999; 990US-0156559.  
28-SEP-1999; 990US-0156458.  
29-SEP-1999; 990US-0156596.  
04-OCT-1999; 990US-0157117.  
05-OCT-1999; 990US-0157753.  
06-OCT-1999; 990US-0157865.  
07-OCT-1999; 990US-0158029.  
08-OCT-1999; 990US-0158323.  
12-OCT-1999; 990US-0158369.  
13-OCT-1999; 990US-0159293.  
13-OCT-1999; 990US-015994.  
13-OCT-1999; 990US-0159229.  
14-OCT-1999; 990US-0159329.  
14-OCT-1999; 990US-0159330.  
14-OCT-1999; 990US-0159331.  
14-OCT-1999; 990US-0159637.  
14-OCT-1999; 990US-0159638.  
18-OCT-1999; 990US-0159584.  
21-OCT-1999; 990US-0160741.  
21-OCT-1999; 990US-0160767.  
21-OCT-1999; 990US-0160770.  
21-OCT-1999; 990US-0160814.  
21-OCT-1999; 990US-0160815.  
22-OCT-1999; 990US-0160980.  
22-OCT-1999; 990US-0160981.  
22-OCT-1999; 990US-0161404.  
25-OCT-1999; 990US-0161409.

PR	08-APR-1999;	99US-0128714.	PR	19-JUL-1999;	99US-0144333.
PR	19-APR-1999;	99US-012945.	PR	19-JUL-1999;	99US-0144334.
PR	21-APR-1999;	99US-013007.	PR	19-JUL-1999;	99US-0144335.
PR	23-APR-1999;	99US-0130449.	PR	20-JUL-1999;	99US-0144352.
PR	23-APR-1999;	99US-0130510.	PR	20-JUL-1999;	99US-0144632.
PR	28-APR-1999;	99US-0130891.	PR	20-JUL-1999;	99US-0144884.
PR	28-APR-1999;	99US-0131449.	PR	21-JUL-1999;	99US-0144814.
PR	30-APR-1999;	99US-0132048.	PR	21-JUL-1999;	99US-0145056.
PR	04-MAY-1999;	99US-0132467.	PR	21-JUL-1999;	99US-0145088.
PR	05-MAY-1999;	99US-0132484.	PR	22-JUL-1999;	99US-0145085.
PR	06-MAY-1999;	99US-0132485.	PR	22-JUL-1999;	99US-0145087.
PR	06-MAY-1999;	99US-0132486.	PR	22-JUL-1999;	99US-0145089.
PR	07-MAY-1999;	99US-0132487.	PR	22-JUL-1999;	99US-0145192.
PR	07-MAY-1999;	99US-0132883.	PR	23-JUL-1999;	99US-0145145.
PR	11-MAY-1999;	99US-0134216.	PR	23-JUL-1999;	99US-0145218.
PR	14-MAY-1999;	99US-0134218.	PR	23-JUL-1999;	99US-0145224.
PR	14-MAY-1999;	99US-0134219.	PR	26-JUL-1999;	99US-0145276.
PR	14-MAY-1999;	99US-0134221.	PR	27-JUL-1999;	99US-0145291.
PR	18-MAY-1999;	99US-0134370.	PR	27-JUL-1999;	99US-0145318.
PR	19-MAY-1999;	99US-0134768.	PR	27-JUL-1999;	99US-0145919.
PR	19-MAY-1999;	99US-0134941.	PR	28-JUL-1999;	99US-0145951.
PR	20-MAY-1999;	99US-0135124.	PR	02-AUG-1999;	99US-0146386.
PR	21-MAY-1999;	99US-0135353.	PR	02-AUG-1999;	99US-0146388.
PR	24-MAY-1999;	99US-0135629.	PR	02-AUG-1999;	99US-0146389.
PR	25-MAY-1999;	99US-0136021.	PR	03-AUG-1999;	99US-0147038.
PR	27-MAY-1999;	99US-0136392.	PR	04-AUG-1999;	99US-0147204.
PR	28-MAY-1999;	99US-0136782.	PR	04-AUG-1999;	99US-0147302.
PR	01-JUN-1999;	99US-0137222.	PR	05-AUG-1999;	99US-0147192.
PR	03-JUN-1999;	99US-0137528.	PR	05-AUG-1999;	99US-0147260.
PR	04-JUN-1999;	99US-0137750.	PR	06-AUG-1999;	99US-0147303.
PR	07-JUN-1999;	99US-0137724.	PR	06-AUG-1999;	99US-0147416.
PR	08-JUN-1999;	99US-0138094.	PR	09-AUG-1999;	99US-0147493.
PR	10-JUN-1999;	99US-0138540.	PR	09-AUG-1999;	99US-0147495.
PR	14-JUN-1999;	99US-0138847.	PR	10-AUG-1999;	99US-0148171.
PR	16-JUN-1999;	99US-0139452.	PR	11-AUG-1999;	99US-0148319.
PR	16-JUN-1999;	99US-0139453.	PR	12-AUG-1999;	99US-0148411.
PR	17-JUN-1999;	99US-0139492.	PR	13-AUG-1999;	99US-0148565.
PR	18-JUN-1999;	99US-0139454.	PR	16-AUG-1999;	99US-0149168.
PR	18-JUN-1999;	99US-0139455.	PR	17-AUG-1999;	99US-0149175.
PR	18-JUN-1999;	99US-0139456.	PR	18-AUG-1999;	99US-014926.
PR	18-JUN-1999;	99US-0139457.	PR	20-AUG-1999;	99US-0149722.
PR	18-JUN-1999;	99US-0139458.	PR	20-AUG-1999;	99US-0149723.
PR	18-JUN-1999;	99US-0139459.	PR	20-AUG-1999;	99US-0149929.
PR	18-JUN-1999;	99US-0139460.	PR	23-AUG-1999;	99US-0149902.
PR	18-JUN-1999;	99US-0139461.	PR	23-AUG-1999;	99US-0149930.
PR	18-JUN-1999;	99US-0139462.	PR	25-AUG-1999;	99US-0150566.
PR	18-JUN-1999;	99US-0139463.	PR	26-AUG-1999;	99US-0150884.
PR	18-JUN-1999;	99US-0139465.	PR	27-AUG-1999;	99US-0151065.
PR	18-JUN-1999;	99US-0139750.	PR	10-SEP-1999;	99US-0153070.
PR	18-JUN-1999;	99US-0139459.	PR	13-SEP-1999;	99US-015375.
PR	21-JUN-1999;	99US-0139763.	PR	27-AUG-1999;	99US-0151080.
PR	21-JUN-1999;	99US-0139817.	PR	15-SEP-1999;	99US-0151103.
PR	22-JUN-1999;	99US-0139899.	PR	30-AUG-1999;	99US-0151138.
PR	23-JUN-1999;	99US-0140353.	PR	01-SEP-1999;	99US-0151940.
PR	23-JUN-1999;	99US-0140354.	PR	07-SEP-1999;	99US-0152363.
PR	24-JUN-1999;	99US-0140695.	PR	10-SEP-1999;	99US-0153070.
PR	28-JUN-1999;	99US-0142055.	PR	13-SEP-1999;	99US-015375.
PR	06-JUL-1999;	99US-0142390.	PR	24-SEP-1999;	99US-0155659.
PR	08-JUL-1999;	99US-0142803.	PR	28-SEP-1999;	99US-0156438.
PR	09-JUL-1999;	99US-0141287.	PR	29-SEP-1999;	99US-0156516.
PR	12-JUL-1999;	99US-0142920.	PR	04-OCT-1999;	99US-0157117.
PR	13-JUL-1999;	99US-0145542.	PR	05-OCT-1999;	99US-0157753.
PR	01-JUL-1999;	99US-0142154.	PR	15-OCT-1999;	99US-0157855.
PR	02-JUL-1999;	99US-0142055.	PR	07-OCT-1999;	99US-0158029.
PR	06-JUL-1999;	99US-0142393.	PR	08-OCT-1999;	99US-0158222.
PR	16-JUL-1999;	99US-0144085.	PR	12-OCT-1999;	99US-0158367.
PR	19-JUL-1999;	99US-0144086.	PR	13-OCT-1999;	99US-0159233.
PR	19-JUL-1999;	99US-0144332.	PR	13-OCT-1999;	99US-0159244.

**RESULT** 11  
 AAG16037  
 CD AAG16037 standard; Protein; 704 AA.  
 XX  
 AC  
 XX  
 AAG16037;  
 17-OCT-2000 (first entry)  
**Arabidopsis thaliana** protein fragment SEQ ID NO: 16523.  
 Protein identification; signal transduction pathway; metabolic pathway;  
 hybridisation assay; genetic mapping; gene expression control; promoter  
 termination sequence



Db 549 efvim 553 | : 0139455 .  
 RESULT 12  
 AAG48654 ID AAG48654 standard; Protein; 704 AA.  
 XX AC AAG48654;  
 XX DT 18-OCT-2000 (first entry)  
 XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 61464 .  
 XX KW Protein identification; signal transduction pathway; metabolic pathway;  
 XX hybridisation assay; genetic mapping; gene expression control; promoter;  
 XX termination sequence.  
 OS Arabidopsis thaliana .  
 XX PN EP1033405-A2.  
 XX PD 06-SEP-2000 .  
 XX PF 25-FEB-2000; 2000EP-0301439 .  
 XX PR 05-FEB-1999; 99US-0121825 .  
 PR 05-MAR-1999; 99US-0123180 .  
 PR 09-MAR-1999; 99US-0123548 .  
 PR 23-MAR-1999; 99US-0115788 .  
 PR 25-MAR-1999; 99US-0126264 .  
 PR 29-MAR-1999; 99US-0126785 .  
 PR 01-APR-1999; 99US-0124462 .  
 PR 06-APR-1999; 99US-0128234 .  
 PR 08-APR-1999; 99US-0128714 .  
 PR 16-APR-1999; 99US-0122845 .  
 PR 19-APR-1999; 99US-0130077 .  
 PR 21-APR-1999; 99US-0110449 .  
 PR 23-APR-1999; 99US-010510 .  
 PR 23-APR-1999; 99US-0130891 .  
 PR 28-APR-1999; 99US-0131449 .  
 PR 30-APR-1999; 99US-01202048 .  
 PR 30-APR-1999; 99US-0132407 .  
 PR 04-MAY-1999; 99US-0134284 .  
 PR 05-MAY-1999; 99US-012485 .  
 PR 06-MAY-1999; 99US-0132486 .  
 PR 06-MAY-1999; 99US-012487 .  
 PR 07-MAY-1999; 99US-0122863 .  
 PR 11-MAY-1999; 99US-0134256 .  
 PR 14-MAY-1999; 99US-0134218 .  
 PR 14-MAY-1999; 99US-0134219 .  
 PR 14-MAY-1999; 99US-0134221 .  
 PR 14-MAY-1999; 99US-0134370 .  
 PR 18-MAY-1999; 99US-0134768 .  
 PR 19-MAY-1999; 99US-0134941 .  
 PR 20-MAY-1999; 99US-0135124 .  
 PR 21-MAY-1999; 99US-0135353 .  
 PR 04-JUN-1999; 99US-0137502 .  
 PR 25-JUN-1999; 99US-0136629 .  
 PR 07-JUN-1999; 99US-0136021 .  
 PR 08-JUN-1999; 99US-013692 .  
 PR 27-MAY-1999; 99US-0136782 .  
 PR 01-JUN-1999; 99US-0137222 .  
 PR 03-JUN-1999; 99US-0137528 .  
 PR 04-JUN-1999; 99US-0137502 .  
 PR 07-JUN-1999; 99US-0137724 .  
 PR 08-JUN-1999; 99US-0138094 .  
 PR 10-JUN-1999; 99US-0138840 .  
 PR 10-JUN-1999; 99US-0138847 .  
 PR 14-JUN-1999; 99US-0139119 .  
 PR 16-JUN-1999; 99US-0139452 .  
 PR 17-JUN-1999; 99US-0139453 .  
 PR 18-JUN-1999; 99US-0139454 .  
 PR 18-JUN-1999; 99US-0139455 .  
 PR 18-JUN-1999; 99US-0139456 .  
 PR 18-JUN-1999; 99US-0139457 .  
 PR 18-JUN-1999; 99US-0139458 .  
 PR 18-JUN-1999; 99US-0139459 .  
 PR 18-JUN-1999; 99US-0139460 .  
 PR 18-JUN-1999; 99US-0139461 .  
 PR 18-JUN-1999; 99US-0139462 .  
 PR 18-JUN-1999; 99US-0139463 .  
 PR 18-JUN-1999; 99US-013970 .  
 PR 18-JUN-1999; 99US-0139763 .  
 PR 21-JUN-1999; 99US-0139817 .  
 PR 22-JUN-1999; 99US-0139891 .  
 PR 23-JUN-1999; 99US-0140353 .  
 PR 23-JUN-1999; 99US-0140354 .  
 PR 24-JUN-1999; 99US-0140693 .  
 PR 28-JUN-1999; 99US-0140823 .  
 PR 29-JUN-1999; 99US-0140991 .  
 PR 30-JUN-1999; 99US-0141287 .  
 PR 01-JUL-1999; 99US-0141812 .  
 PR 01-JUL-1999; 99US-0142154 .  
 PR 02-JUL-1999; 99US-0142055 .  
 PR 06-JUL-1999; 99US-0142390 .  
 PR 08-JUL-1999; 99US-0142803 .  
 PR 09-JUL-1999; 99US-0142920 .  
 PR 12-JUL-1999; 99US-0142977 .  
 PR 13-JUL-1999; 99US-0143562 .  
 PR 14-JUL-1999; 99US-0143624 .  
 PR 15-JUL-1999; 99US-0144005 .  
 PR 16-JUL-1999; 99US-0144085 .  
 PR 17-JUL-1999; 99US-0144086 .  
 PR 19-JUL-1999; 99US-0144325 .  
 PR 19-JUL-1999; 99US-0144331 .  
 PR 19-JUL-1999; 99US-0144332 .  
 PR 19-JUL-1999; 99US-0144333 .  
 PR 19-JUL-1999; 99US-0144334 .  
 PR 19-JUL-1999; 99US-0144335 .  
 PR 20-JUL-1999; 99US-0144352 .  
 PR 20-JUL-1999; 99US-0144632 .  
 PR 20-JUL-1999; 99US-0144884 .  
 PR 21-JUL-1999; 99US-0144814 .  
 PR 21-JUL-1999; 99US-0145086 .  
 PR 21-JUL-1999; 99US-0145088 .  
 PR 22-JUL-1999; 99US-0145085 .  
 PR 22-JUL-1999; 99US-0145087 .  
 PR 22-JUL-1999; 99US-0145089 .  
 PR 22-JUL-1999; 99US-0145192 .  
 PR 23-JUL-1999; 99US-0145145 .  
 PR 23-JUL-1999; 99US-0145218 .  
 PR 26-JUL-1999; 99US-0145244 .  
 PR 27-JUL-1999; 99US-0145276 .  
 PR 27-JUL-1999; 99US-0145913 .  
 PR 27-JUL-1999; 99US-0145918 .  
 PR 27-JUL-1999; 99US-0145919 .  
 PR 28-JUL-1999; 99US-0147302 .  
 PR 02-AUG-1999; 99US-0146386 .  
 PR 02-AUG-1999; 99US-0146388 .  
 PR 02-AUG-1999; 99US-0146389 .  
 PR 03-AUG-1999; 99US-0147038 .  
 PR 04-AUG-1999; 99US-0147204 .  
 PR 04-AUG-1999; 99US-0147493 .  
 PR 09-AUG-1999; 99US-0147935 .  
 PR 10-AUG-1999; 99US-0147192 .  
 PR 05-AUG-1999; 99US-0147260 .  
 PR 11-AUG-1999; 99US-0148319 .  
 PR 12-AUG-1999; 99US-0148341 .  
 PR 13-AUG-1999; 99US-0148565 .  
 PR 13-AUG-1999; 99US-0148684 .  
 PR 16-AUG-1999; 99US-0149368 .

17-AUG-1999; PR	99US-0149175;
18-AUG-1999; PR	99US-0149122;
20-AUG-1999; PR	99US-0149123;
20-AUG-1999; PR	99US-0149029;
23-AUG-1999; PR	99US-0149029;
23-AUG-1999; PR	99US-0149030;
25-AUG-1999; PR	99US-0150666;
26-AUG-1999; PR	99US-0150884;
01-SEP-1999; PR	99US-0152633;
10-SEP-1999; PR	99US-0153070;
13-SEP-1999; PR	99US-0153758;
15-SEP-1999; PR	99US-0154018;
16-SEP-1999; PR	99US-0154139;
20-SEP-1999; PR	99US-0154779;
22-SEP-1999; PR	99US-0155139;
23-SEP-1999; PR	99US-0155486;
24-SEP-1999; PR	99US-0155659;
28-SEP-1999; PR	99US-0156158;
29-SEP-1999; PR	99US-0156596;
04-OCT-1999; PR	99US-0157117;
05-OCT-1999; PR	99US-0157753;
06-OCT-1999; PR	99US-0157965;
07-OCT-1999; PR	99US-0158029;
08-OCT-1999; PR	99US-0158222;
12-OCT-1999; PR	99US-0158369;
13-OCT-1999; PR	99US-0159293;
13-OCT-1999; PR	99US-0159294;
13-OCT-1999; PR	99US-0159295;
14-OCT-1999; PR	99US-0159329;
14-OCT-1999; PR	99US-0159331;
14-OCT-1999; PR	99US-0159637;
14-OCT-1999; PR	99US-0159638;
18-OCT-1999; PR	99US-0159594;
21-OCT-1999; PR	99US-0160741;
21-OCT-1999; PR	99US-0160742;
22-OCT-1999; PR	99US-0160980;
22-OCT-1999; PR	99US-0160981;
22-OCT-1999; PR	99US-0160982;
21-OCT-1999; PR	99US-0160768;
21-OCT-1999; PR	99US-0160814;
21-OCT-1999; PR	99US-0160815;
22-OCT-1999; PR	99US-0161359;
26-OCT-1999; PR	99US-0161360;
26-OCT-1999; PR	99US-0161361;
28-OCT-1999; PR	99US-0161920;
28-OCT-1999; PR	99US-0161992;
28-OCT-1999; PR	99US-0161993;
29-OCT-1999; PR	99US-0162142.

Qy	123	IIPAIL-----KDNQEYELPADGEAMLFSAHLNSAIKP.KPLFLHKAQREERY	174
Db	348	rhlgyllknccsvrcrkldngtlleds11----cvikgtg-----kyrrswrydl	394
Qy	175	IS-----YLLG--LRELRKAIAKV-----EOKVTLKAVKDIEAVP--VVAV	213
Db	395	vdsf1ppstgellavaldervlgfirnavsvffstmedgki----vkglagprdkggpvly	450
Qy	214	NTAYMLGIGRPLPNKPKVASYIEDKDNTLL-----YGTDTBPIGY	254
Db	451	gyhmlmglegpme----aflek-nlifrgmahpvlsvndpakafdygdwkvfg-	503
Qy	255	RDIAGYRMSEGVLELVIDLNSELCLPS---ELKHSCELYRTSSWAPDKSLRIVREDE	311
Db	504	---ayptatnlkxldkskhv1lfpggarealhngey-----kiwpeqq	548
Qy	312	GNARL	316
Db	549	efvrn	553
RESULT 13			
AAG16036 standard; Protein; 726 AA.			
XX	AAG16036;		
XX	AC		
XX	DT	17-OCT-2000	(first entry)
XX	DE	Arabidopsis thaliana protein fragment SEQ ID NO: 16522.	
KW	protein identification; signal transduction pathway; metabolic pathway		
KW	hybridisation assay; genetic mapping; gene expression control; promotion		
KW	termination sequence.		
XX	OS	Arabidopsis thaliana.	
XX	PN	EP103405-A2.	
XX	PD	06-SEP-2000.	
XX	PF	25-FEB-2000; 2000EP-0301439.	
XX	PR	25-FEB-1999; 99US-0121825.	
PR	05-MAR-1999;	99US-0123180.	
PR	09-MAR-1999;	99US-0123348.	
PR	23-MAR-1999;	99US-0125788	
PR	25-MAR-1999;	99US-0126264.	
PR	29-MAR-1999;	99US-0126785.	
PR	01-APR-1999;	99US-0127462.	
PR	06-APR-1999;	99US-0128234.	
PR	08-APR-1999;	99US-0128714.	
PR	16-APR-1999;	99US-0129845.	
PR	19-APR-1999;	99US-0130077.	
PR	21-APR-1999;	99US-0130449.	
PR	23-APR-1999;	99US-0130510	
PR	23-APR-1999;	99US-0130891.	
PR	28-APR-1999;	99US-0131449.	
PR	30-APR-1999;	99US-0132048	
PR	30-APR-1999;	99US-0132407.	
PR	04-MAY-1999;	99US-0132454.	
PR	05-MAY-1999;	99US-0132485.	
PR	06-MAY-1999;	99US-0132486.	
PR	06-MAY-1999;	99US-0132447.	
PR	07-MAY-1999;	99US-0132863	
PR	11-MAY-1999;	99US-0134556.	
PR	14-MAY-1999;	99US-0134218.	
PR	14-MAY-1999;	99US-0134219.	
PR	14-MAY-1999;	99US-0134221.	
PR	14-MAY-1999;	99US-0134370.	
PR	18-MAY-1999;	99US-0134768.	
PR	19-MAY-1999;	99US-0134941.	
PR	20-MAY-1999;	99US-0135124.	

21-MAY-1999; 99US-0135353.  
 PR 24-MAY-1999; 99US-0135629.  
 PR 25-MAY-1999; 99US-0135021.  
 PR 27-MAY-1999; 99US-0136392.  
 PR 28-MAY-1999; 99US-0136782.  
 PR 01-JUN-1999; 99US-013222.  
 PR 03-JUN-1999; 99US-0131528.  
 PR 04-JUN-1999; 99US-0131502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138904.  
 PR 10-JUN-1999; 99US-0138450.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 14-JUN-1999; 99US-0138119.  
 PR 16-JUN-1999; 99US-013452.  
 PR 16-JUN-1999; 99US-0134453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0134455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-013452.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-013461.  
 PR 18-JUN-1999; 99US-013462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139470.  
 PR 18-JUN-1999; 99US-0139475.  
 PR 18-JUN-1999; 99US-0139476.  
 PR 18-JUN-1999; 99US-0139476.  
 PR 21-JUN-1999; 99US-0138817.  
 PR 22-JUN-1999; 99US-0138899.  
 PR 22-JUN-1999; 99US-014353.  
 PR 23-JUN-1999; 99US-014354.  
 PR 24-JUN-1999; 99US-0140695.  
 PR 28-JUN-1999; 99US-0144823.  
 PR 29-JUN-1999; 99US-014091.  
 PR 30-JUN-1999; 99US-014287.  
 PR 01-JUL-1999; 99US-0141842.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0144055.  
 PR 06-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0144903.  
 PR 09-JUL-1999; 99US-014220.  
 PR 12-JUL-1999; 99US-0143977.  
 PR 13-JUL-1999; 99US-0143442.  
 PR 14-JUL-1999; 99US-0143624.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144335.  
 PR 19-JUL-1999; 99US-0144325.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144452.  
 PR 20-JUL-1999; 99US-0144622.  
 PR 22-JUL-1999; 99US-0144884.  
 PR 22-JUL-1999; 99US-0145019.  
 PR 22-JUL-1999; 99US-0145192.  
 PR 23-JUL-1999; 99US-0145145.  
 PR 23-JUL-1999; 99US-0145086.  
 PR 23-JUL-1999; 99US-0145088.  
 PR 23-JUL-1999; 99US-0145085.  
 PR 22-JUL-1999; 99US-0145087.  
 PR 26-JUL-1999; 99US-0145226.  
 PR 27-JUL-1999; 99US-0145913.  
 PR 27-JUL-1999; 99US-0145918.  
 PR 28-JUL-1999; 99US-0145919.  
 PR 28-JUL-1999; 99US-0145551.  
 PR 02-AUG-1999; 99US-0146386.  
 PR 02-AUG-1999; 99US-0146388.  
 PR 02-AUG-1999; 99US-0146389.  
 PR 03-AUG-1999; 99US-0147038.  
 PR 04-AUG-1999; 99US-0147204.  
 PR 04-AUG-1999; 99US-0147302.  
 PR 05-AUG-1999; 99US-0147192.  
 PR 05-AUG-1999; 99US-0147260.  
 PR 06-AUG-1999; 99US-0147303.  
 PR 06-AUG-1999; 99US-0147416.  
 PR 07-AUG-1999; 99US-0148565.  
 PR 07-AUG-1999; 99US-0147493.  
 PR 09-AUG-1999; 99US-0147935.  
 PR 10-AUG-1999; 99US-0149175.  
 PR 11-AUG-1999; 99US-0148319.  
 PR 12-AUG-1999; 99US-0149723.  
 PR 13-AUG-1999; 99US-0149723.  
 PR 13-AUG-1999; 99US-0149929.  
 PR 16-AUG-1999; 99US-0149938.  
 PR 17-AUG-1999; 99US-0149939.  
 PR 18-AUG-1999; 99US-0149426.  
 PR 20-AUG-1999; 99US-0149722.  
 PR 20-AUG-1999; 99US-0149723.  
 PR 23-AUG-1999; 99US-0149902.  
 PR 23-AUG-1999; 99US-0149930.  
 PR 25-AUG-1999; 99US-0150566.  
 PR 26-AUG-1999; 99US-0150884.  
 PR 27-AUG-1999; 99US-0151065.  
 PR 27-AUG-1999; 99US-0151066.  
 PR 27-AUG-1999; 99US-0151080.  
 PR 30-AUG-1999; 99US-0151303.  
 PR 31-AUG-1999; 99US-0151438.  
 PR 01-SEP-1999; 99US-0151930.  
 PR 07-SEP-1999; 99US-0152363.  
 PR 10-SEP-1999; 99US-0153070.  
 PR 13-SEP-1999; 99US-0153758.  
 PR 15-SEP-1999; 99US-0154018.  
 PR 16-SEP-1999; 99US-014039.  
 PR 20-SEP-1999; 99US-0154779.  
 PR 22-SEP-1999; 99US-0155139.  
 PR 23-SEP-1999; 99US-0155753.  
 PR 24-SEP-1999; 99US-0155486.  
 PR 28-SEP-1999; 99US-0155659.  
 PR 29-SEP-1999; 99US-0155956.  
 PR 04-OCT-1999; 99US-0151117.  
 PR 05-OCT-1999; 99US-0157753.  
 PR 06-OCT-1999; 99US-0157865.  
 PR 07-OCT-1999; 99US-0156029.  
 PR 08-OCT-1999; 99US-0158232.  
 PR 12-OCT-1999; 99US-0158369.  
 PR 13-OCT-1999; 99US-0159293.  
 PR 13-OCT-1999; 99US-0159294.  
 PR 14-OCT-1999; 99US-0159295.  
 PR 14-OCT-1999; 99US-0159329.  
 PR 14-OCT-1999; 99US-0159330.  
 PR 14-OCT-1999; 99US-0159331.  
 PR 21-OCT-1999; 99US-0160768.  
 PR 21-OCT-1999; 99US-0160770.  
 PR 21-OCT-1999; 99US-0160811.  
 PR 21-OCT-1999; 99US-0160815.  
 PR 22-OCT-1999; 99US-0160880.  
 PR 22-OCT-1999; 99US-0160891.  
 PR 22-OCT-1999; 99US-0160899.  
 PR 25-OCT-1999; 99US-0161404.  
 PR 25-OCT-1999; 99US-0161405.  
 PR 25-OCT-1999; 99US-0161406.  
 PR 26-OCT-1999; 99US-0161359.  
 PR 26-OCT-1999; 99US-0161360.  
 PR 26-OCT-1999; 99US-0161361.

PR 28-OCT-1999; 99US-0161920.  
 PR 28-OCT-1999; 99US-0161923.  
 PR 29-OCT-1999; 99US-0161993.  
 PR 29-OCT-1999; 99US-0162142.

Query Match Similarity 5.5%; Score 104; DB 21; Length 726;  
 Best Local Similarity 22.5%; Pred. No. 0 23; Mismatches 49; Gaps 20;

Matches 82; Conservative 49; Indels 122; Gaps 20;

Qy 29 IPIVETLKEEFPGLNITGTYKFLPKDKTIDLVGGIASDL---IEIIGTSYTHAIL 83  
 Db 256 ipie-nvpee,hf---tvpaisflmgdpkmatlgqngqptgvkrekirgrktm 311  
 Qy 84 PILLPSRVEAQWORD-----REVKEELFEL-SPKGFWLPPELAYDP 122  
 Db 312 pl-lseggipretlwkkilrsgcayansrihavqevlyasgdmnlpseqeak 369  
 Qy 123 ITAIL-----XONGYEFLFADGEAMLESAHNSAKPIKLYPHLIKAREKRRY 174  
 Db 370 rhhgllknkncsvrofkgntllleesisl-----tvikqtg-----kyrrswrydl 416  
 Qy 175 IS-----YLGS-LREPLRAIKLVF---EGKVTLKAVKDKTEAVP--VVWAV 213  
 :| 417 vsdfllpskgelayaldevflrnavaqsfstmedqk1----vkqlaqvpdkgvly 472  
 Qy 214 NTAVMLGIGRLPLMNPKAWSIEDKDNL1-----YGDIEFFGY 254  
 :| 473 grymlmgielgpmse-----afikek-nlfrgmahpvlsvndnpakafdygdkvkg- 525  
 Qy .255 RDIAGYNSVEGLIVDELNSKLEIPS--ELKSYREYLRTSWSADKSRIWRDDE 311  
 Db 536 ---ayptatnlkfldskshvlfpggarealmrgedy-----kliwpeqq 570  
 Qy 31.2 GNARL 316  
 Db 571 efvrm 575

RESULT 14  
 AAG8653 standard; Protein; 726 AA.  
 XX AC AAG8653;  
 XX ID AAG8653 standard; Protein; 726 AA.  
 XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 61463.  
 XX DT 18-OCT-2000 (first entry)  
 XX DE Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX ARABIDOPSIS THALIANA  
 OS Arabidopsis thaliana.  
 XX PN EP1033405-A2.  
 XX 06-SEP-2000.  
 XX 25-FEB-2000; 2000EP-0101439.  
 XX PR 25-FEB-1999; 99US-0121825.  
 PR 05-MAR-1999; 99US-0123180.  
 PR 09-MAR-1999; 99US-0123548.  
 PR 23-MAR-1999; 99US-0125788.  
 PR 25-MAR-1999; 99US-0126264.  
 PR 29-MAR-1999; 99US-0126785.  
 PR 01-APR-1999; 99US-0127462.  
 PR 06-APR-1999; 99US-0128234.  
 PR 08-APR-1999; 99US-0128714.  
 PR 16-APR-1999; 99US-0129845.  
 PR 19-APR-1999; 99US-0130077.  
 PR 21-APR-1999; 99US-0130449.  
 PR 23-APR-1999; 99US-0130510.  
 PR 23-APR-1999; 99US-0130891.  
 PR 28-APR-1999; 99US-0131449.  
 PR 30-APR-1999; 99US-0132048.  
 PR 04-MAY-1999; 99US-0132407.  
 PR 05-MAY-1999; 99US-0132485.  
 PR 06-MAY-1999; 99US-0132486.  
 PR 06-MAY-1999; 99US-0132487.  
 PR 07-MAY-1999; 99US-0132463.  
 PR 11-MAY-1999; 99US-0134256.  
 PR 14-MAY-1999; 99US-0134218.  
 PR 14-MAY-1999; 99US-013419.  
 PR 14-MAY-1999; 99US-0134221.  
 PR 18-MAY-1999; 99US-0134768.  
 PR 19-MAY-1999; 99US-0134769.  
 PR 20-MAY-1999; 99US-0135124.  
 PR 21-MAY-1999; 99US-0135153.  
 PR 24-MAY-1999; 99US-0135639.  
 PR 25-MAY-1999; 99US-0136021.  
 PR 27-MAY-1999; 99US-0136592.  
 PR 01-JUN-1999; 99US-0136782.  
 PR 02-JUN-1999; 99US-0137222.  
 PR 03-JUN-1999; 99US-0137328.  
 PR 04-JUN-1999; 99US-0137502.  
 PR 07-JUN-1999; 99US-0137724.  
 PR 08-JUN-1999; 99US-0138054.  
 PR 10-JUN-1999; 99US-0138400.  
 PR 10-JUN-1999; 99US-0138847.  
 PR 11-JUN-1999; 99US-0139119.  
 PR 16-JUN-1999; 99US-0139452.  
 PR 16-JUN-1999; 99US-0139453.  
 PR 17-JUN-1999; 99US-0139492.  
 PR 18-JUN-1999; 99US-0139454.  
 PR 18-JUN-1999; 99US-0139455.  
 PR 18-JUN-1999; 99US-0139456.  
 PR 18-JUN-1999; 99US-0139457.  
 PR 18-JUN-1999; 99US-0139458.  
 PR 18-JUN-1999; 99US-0139459.  
 PR 18-JUN-1999; 99US-0139460.  
 PR 18-JUN-1999; 99US-0139461.  
 PR 18-JUN-1999; 99US-0139462.  
 PR 18-JUN-1999; 99US-0139463.  
 PR 18-JUN-1999; 99US-0139470.  
 PR 18-JUN-1999; 99US-0139476.  
 PR 21-JUN-1999; 99US-0139817.  
 PR 22-JUN-1999; 99US-0139818.  
 PR 23-JUN-1999; 99US-0140333.  
 PR 23-JUN-1999; 99US-0140544.  
 PR 24-JUN-1999; 99US-0140995.  
 PR 28-JUN-1999; 99US-0140833.  
 PR 29-JUN-1999; 99US-0140991.  
 PR 30-JUN-1999; 99US-0141287.  
 PR 01-JUL-1999; 99US-0141822.  
 PR 01-JUL-1999; 99US-0142154.  
 PR 02-JUL-1999; 99US-0142055.  
 PR 05-JUL-1999; 99US-0142390.  
 PR 08-JUL-1999; 99US-0142813.  
 PR 09-JUL-1999; 99US-0142920.  
 PR 12-JUL-1999; 99US-0142977.  
 PR 13-JUL-1999; 99US-0143592.  
 PR 14-JUL-1999; 99US-0143644.  
 PR 15-JUL-1999; 99US-0144005.  
 PR 16-JUL-1999; 99US-0144085.  
 PR 16-JUL-1999; 99US-0144086.  
 PR 19-JUL-1999; 99US-0144225.  
 PR 19-JUL-1999; 99US-0144331.  
 PR 19-JUL-1999; 99US-0144332.  
 PR 19-JUL-1999; 99US-0144333.  
 PR 19-JUL-1999; 99US-0144334.  
 PR 19-JUL-1999; 99US-0144335.  
 PR 20-JUL-1999; 99US-0144336.  
 PR 20-JUL-1999; 99US-0144632.

PR	14-OCT-1999;	99US-01595638;
PR	18-OCT-1999;	99US-01595749;
PR	21-OCT-1999;	99US-01607411;
PR	21-OCT-1999;	99US-0160747;
PR	21-OCT-1999;	99US-0160768;
PR	21-OCT-1999;	99US-0160770;
PR	21-OCT-1999;	99US-0160814;
PR	21-OCT-1999;	99US-0160815;
PR	22-OCT-1999;	99US-0160940;
PR	22-OCT-1999;	99US-0160981;
PR	22-OCT-1999;	99US-0161099;
PR	25-OCT-1999;	99US-0161404;
PR	25-OCT-1999;	99US-0161405;
PR	25-OCT-1999;	99US-0161406;
PR	26-OCT-1999;	99US-0161359;
PR	26-OCT-1999;	99US-0161360;
PR	28-OCT-1999;	99US-0161920;
PR	28-OCT-1999;	99US-0161992;
PR	28-OCT-1999;	99US-0161993;
PR	29-OCT-1999;	99US-0162142.
Query Match 5.5%;		
Best Local Similarity 22.5%;		
Matches 82; Conservative 4		
Qy	29 IPVIELIKEIPFGLNITGTYE	
Db	256 1pile-mpeelh--tvpval	
Qy	84 PILPSRVEAQYRD-----	
Db	312 p1--l-seigggipretllwkkl	
Qy	123 IIPAIL-----KDNGYEYL	
Db	370 ringllknksvcrfkanghtlli	
Qy	175 IS-----YLIG--LRLIE	
Db	417 vsdfplppskgelayaaldevlgfri	
Qy	214 NTAVMIGIGRPLMNPKKVASW	
Db	473 gyhmunglelpmse-----afii	
Qy	255 RDIAGYRMSSVEELVIDELNSE	
Db	526 ---ayvtatnlkfildkskhw	
Qy	312 GNARL 316	
Db	571 efvrm 575	
RESULT 15		
ABG18270	ID ABG18270 standard; Protein;	
XX	AC ABG18270;	
XX	DT 18-FEB-2002 (first entry)	
XX	DE Novel human diagnostic protein	
KW	Human; chromosome mapping; food supplement; medical imaging	
KW	OS Homo sapiens.	
XX	PN WO200175067-A2.	
XX	PD 11-OCT-2001.	

XX  
PF 30-MAR-2001; 2001WO-US08631.  
XX  
PR 31-MAR-2000; 2000US-0540217.  
PR 23-AUG-2000; 2000US-0649167.  
XX  
PA (HYSE-) HYSEQ INC.  
XX  
PI Drmanac RT, Liu C, Tang YT;  
XX  
WPI; 2001-639362/73.  
DR N-PDB; AASB2457.

XX  
PT New isolated polynucleotide and encoded polypeptides, useful in diagnostics, forensics, gene mapping, identification of mutations responsible for genetic disorders or other traits and to assess PT biodiversity

XX  
PS Claim 20: SEQ ID No 48629; 103pp; English.  
XX  
CC The invention relates to isolated polynucleotide (I) and polypeptide (II) sequences. (I) is useful as hybridisation probes, polymerase chain reaction (PCR) primers, oligomers, and for chromosome CC and gene mapping, and in recombinant production of (II). The CC polynucleotides are also used in diagnostics as expressed sequence tags CC for identifying expressed genes. (I) is useful in gene therapy techniques CC to restore normal activity of (II) or to treat disease states involving CC (II). (II) is useful for generating antibodies against it, detecting or CC quantitating a polypeptide in tissue, as molecular weight markers and as CC a food supplement. (II) and its binding partners are useful in medical CC imaging of sites expressing (II). (I) and (III) are useful for treating CC disorders involving aberrant protein expression or biological activity. CC The polypeptide and polynucleotide sequences have applications in CC diagnostics, forensics, gene mapping, identification of mutations CC responsible for genetic disorders or other traits to assess biodiversity CC and to produce other types of data and products dependent on DNA and CC amino acid sequences. ABG00010-ABG3077 represent novel human CC diagnostic amino acid sequences of the invention.  
CC Note: The sequence data for this patent did not appear in the printed CC specification, but was obtained in electronic format directly from WIPO CC at ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 936 AA;

Query Match Score 103.5; DB 22; Length 936;  
Best Local Similarity 5.5%; Pred. No. 0.39; Mismatches 49; Indels 109; Gaps 19;

Matches 70; Conservative 49; MisMatches 84; Indels 109; Gaps 19;  
Qy 127 IJKDNGYEWLFADGEAMESAHLNSAIKPKLYPHILKAOREK-----RFFRYISYL 178  
Db 183 Ileedngw---lae eally---veseervenglptvpanswlsbrekvaafagaepiylnht 235  
Qy 179 IGL-----RELAIAKIVFEGRT---LLKAV-----KDIEAVPV---W- 210  
Db 236 1gfpvglreilkqdsywagnstreealargrelarhwddqqkqagidlpvgdfawy 295  
Qy 211 -----VAVNTAVMLIGIGRLPLMNPKKVA---SMIEDKDNLIL 244  
Db 296 dhvittsllgnvvarhpkdgssdidlfirgraptgeaaemtkw----- 346  
Qy 245 YGTIDIEFIGYRDIGS-YRMSVEGLLEVDELINSCLPSELKHSGRELYLRTSSWAPDK 302  
Db 347 fntnhympfefvgqgkkitwtcllddeala-----lghkkrpv1gpyvw---- 395  
Qy 303 SIRIMR- EDEGARLMWLSYMRGEALLAENS DARG- W- EP----IPERRLD 348  
Db 396 ---lwlgkvkgeqfdrlsldndilpyqyqqlaeia-krqiewvqidspaivlelpipwid 451  
Qy 349 AFRAIYNDWRGE 360  
Db 452 aykpaydal99 463

Sun Jun 2 18:27:15 2002

us-09-886-400-4.rag

Page 17